Development of a Multiple Linear Regression Model to Predict Non-Working Days Due to Adverse Weather Conditions for a Four-lane, Divided Highway Construction Project in Brazos County, Texas

Rajesh Sake, MSCM, Paul Woods, D.E.D., RA, Michael Speed, Ph.D. and Wonkee Lee, MSCM
Texas A&M University
College Station, Texas

The objective of this research was to develop a statistical model to predict number of non-working days in any month due to adverse weather conditions. A simple linear regression model was developed using the sample data collected from on-going highway construction project on SH 6, a four-lane, divided highway. The significant variables were selected by backward elimination method. With adjusted R square value of 0.719, the model provides a relatively easy and powerful method to predict number of non-working days and hence the project duration. The weather data required for the analysis is easily downloadable from NCDC (National Climatic Data Center) website.

Keywords: Non-working days, Linear Regression, Backward Elimination, Project Duration

Introduction

In the United States, more than 10,000 companies provide highway and street construction services with combined annual revenues of around $70 billion. These companies acquire almost all of their business by bidding on fixed-cost contracts. The duration of highway construction projects is more critical today as there is an increasing number of resurfacing, restoration, and rehabilitation type projects being constructed under traffic, resulting in an increase in the exposure of construction workers and motorists. Traffic volumes on most highways are significantly greater and are continually increasing, thereby creating a greater impact on the motoring public in both safety considerations and cost (D&B, 2008).

Weather is one of the most common causes for delay in construction projects. Due to continuous exposure to an outdoor environment, highway construction projects are greatly affected by weather conditions (Lee, et.al, 2008). But, US General Services Administration clearly states in its instructions to contractors that “Weather which hinders or prevents work is not a basis for a time extension unless it surpasses in severity the weather reasonably to be expected in the locality at the particular time of year”. Texas Department of Transportation (TXDOT) specifies contractors are required to decide project completion dates based on their own risks. This delivery method does not allow for contractors to request time extension due to the adverse weather conditions except for some cataclysmic cases (Kenner, et al., 1998). As a result, significant time and effort are spent in settling disputes between the owner and contractor regarding the reasonable number of non-working days (Woods, et al., 2006). Hence, when adverse weather conditions, which cause non-working days, are not adequately considered in highway project scheduling, delay in the construction project is unavoidable (Kenner, et al., 1998). The delay leads to request for contract time extension.

Although several models have been developed to predict the weather-related, non-working days in a construction project, those methods are not easy to apply and/or are not handy to use for contractors (Lee, et.al, 2008). This study was designed to develop an effective and efficient regression model for engineers and contractors to estimate the number of non-working days for any future construction project in Brazos County provided the start date and project duration are known.

The Problem and Its Setting

Problem Statement

The objective of this research is to develop a linear regression model predicting number of non-working days due to adverse weather conditions for four-lane, divided highway construction projects in Brazos County, Texas.

Assumptions

1) The sample project is a representative of the population.
2) The daily work reports available from TXDOT are complete and accurate.
3) The Local Weather Data available from NCDC used in this research accurately reflects the weather conditions at the construction project site.

Research Objectives

1) To identify weather factors those affect the number of non-working days.
2) To develop a linear regression model to predict the number of non-work days for four-lane, divided highway construction projects in Brazos County, Texas based on weather conditions as recorded at Easter wood Airport and actual non-working days as reported on the daily work reports of a Texas Department of Transportation (TXDOT) project in Brazos County.

Research Method

Data Collection

The data for this research was collected from Texas Department of Transportation (TX DOT) for the highway project which is currently in progress on SH-6 in Brazos County, Texas.

The details of the project are as follows –

- Location – SH-6, Brazos County, Texas
- Length – 6.502 Miles
- Contract Price - $ 100,313,270.79
- Contractor – T.J. Lambert Construction Inc
- Project Start Date – 03/06/2006
- Percentage of Work Completed as of the end of September 2008 – 75%

Daily work reports (DWR) were collected for 30 months, i.e. from April 2006 through September 2008. By examining Daily Work Reports closely, the number of non-working days in each month was recorded for the 30-month period. The recorded data of actual non-working days was used as a dependent variable in the regression analysis and the predicted values calculated from the regression equation were validated by comparing it to the actual number of non-working days as recorded on FM-158 Phase II highway project.

The local weather data was collected from the Easter wood Field Airport weather station, College Station, Texas which is the closest weather station to the project site. All historical weather data used in this research is available to download on National Climatic Data Center (NCDC), National Oceanic and Atmospheric Administration (NOAA) website.

Analysis

Variables

A linear regression model was developed with all the variables which are expected to have a significant influence on number of non-working days as learned from the literature review.

Four independent variables were considered in the process which included, monthly weather factors; Precipitation (P – in mm), Wind Speed (WS in kmph), Temperature (T in degree Centigrade) and the fourth independent variable was calculated as Weekdays and holidays for the month minus eight (i.e. WDH = Weekend days+ Holidays - 8). The reason for subtracting eight from the number of weekend days and holidays per month was to meet the statistical assumptions for a multiple regression, a simple method of data transformation. It can be thought of logically as accounting for the fact that there is no month that has less than a total of eight weekend days and holidays (Woods et. al, 2006). Number of non-working days (NWD) was used as the dependent variable with actual number of non working days for the month as collected from daily work reports of the highway construction project.

Model Considered

The following model was considered to develop a linear regression model –

$$NWD = \beta_0 + \beta_1 (WDH) + \beta_2 (P) + \beta_3 (WS) + \beta_4 (T) + \varepsilon$$

Where, NWD – Number of non-working days, $\beta_0,\beta_1,\beta_2,\beta_3,\beta_4$ are the parameters, WDH = Weekend days + Holidays – 8, P – Precipitation in mm, WS – Wind speed in kmph, T – Temperature in ^{o}C.

Variable selection procedure

The variable selection procedure used in this research is *Backward Elimination method.* This procedure begins with a model which includes all the variables considered and then attempts to delete one variable at a time by determining whether the p-value of the least significant variable is less than the specified or default vale. Once a variable is removed from the model it cannot re-enter the model in the next step. In this method as it works its way down instead of up, the solution with the greatest R-square value is always retained.

Results

Predictive Statistics

Table below shows the results of the regression analysis. The results of analysis of variance (ANOVA) show that the variables, WDH, precipitation and temperature are significant at p-value <=0.05. This means that all these three independent variables are significant predictors of the dependent variable, Number of non-working days. The fourth independent variable, wind speed was eliminated by backward elimination method. The final model developed can be represented as follows –

$$NWD = 11.799 + 1.468\,(WDH) + 0.033\,(P) - 0.148\,(T)$$

The adjusted R-squared for the final model is 0.748 which means that almost 75% of the variability in the variable of interest, number of non-working days can be explained by the final model with the above three variables.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.865[a]	.748	.719	1.77033

a. Predictors: (Constant), Precipitation in mm, Weekend days + Holidays -8, Temperature in degree centigrade

Table 1 : Regression Model Summary

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	241.981	3	80.660	25.737	.000[a]
	Residual	81.486	26	3.134		
	Total	323.467	29			

a. Predictors: (Constant), Precipitation in mm, Weekend days + Holidays -8, Temperature in degree centigrade

b. Dependent Variable: Non-working Days

Table 2 : Results of ANOVA

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	11.799	1.359		8.682	.000
	Weekend days + Holidays -8	1.468	.306	.478	4.796	.000
	Temperature in degree centigrade	-.148	.050	-.298	-2.987	.006
	Precipitation in mm	.033	.005	.629	6.382	.000

a. Dependent Variable: Non-working Days

Table 3 : Coefficients of regression model

Diagnostics

Normality of Residuals

Un-standardized residuals were calculated and the normality of it was tested using Shapiro-Wilks test as the degrees of freedom was less than 50. As the p-value was of 0.616 (>0.05), we failed to reject the null hypothesis that the residuals are normally distributed.

Constant Variance

As the residuals were normal, the constant variance was tested using Breusch-Pagan test for Heteroscedasticity. We failed to reject the null hypothesis of constant variance as the significance value of the test was 0.8212 (>0.05).

Multi-co linearity

The multicollinearity between the independent variables was tested by finding the VIF (Variance Inflation Factor) as shown in the figure below. As a thumb rule, since VIF for all the three independent variables is less than 10 there is no multicollinearity problem in the model.

Validation / Comparison

Comparison of actual and predicted values for FM-158 Phase II highway project

The regression model developed can be validated by comparing it to the actual number of non-working days with the predicted number of non-working days for FM-158 highway construction project for the period from August 2002 to July 2004. The data analyzed by Woods et.al in 2006 was used for the comparison. The following graph shows the comparison of predicted and actual number of working days –

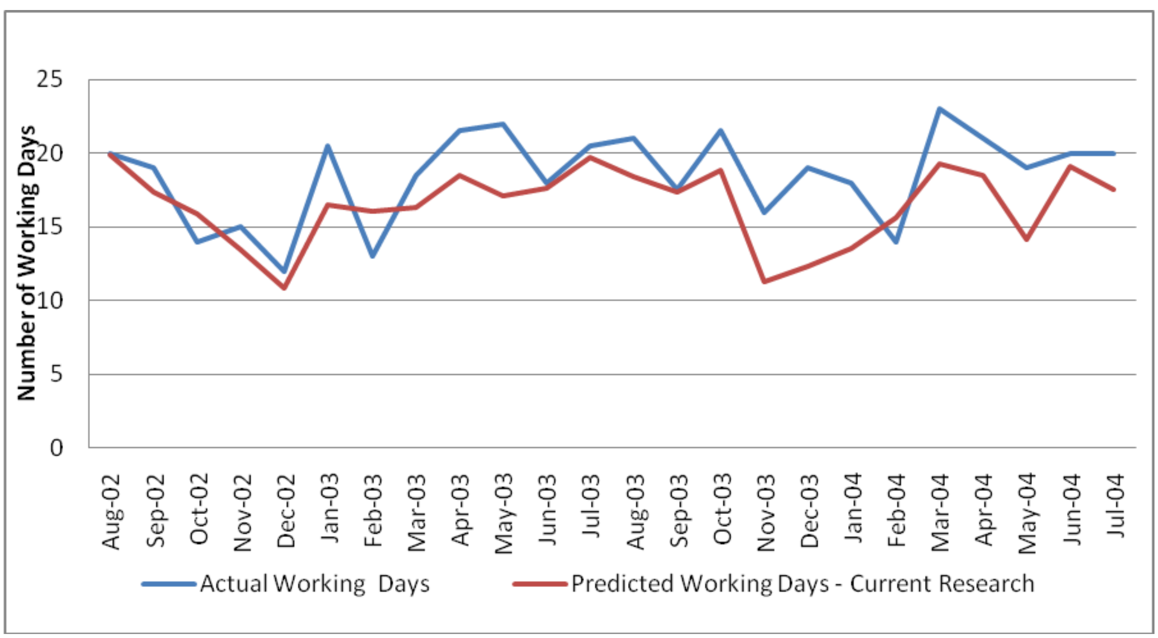

Figure 1 : Comparison of actual and predicted values for FM-158 Project

The comparison of actual and predicted number of days for FM-158 project shows a variation of 11% to the lower side of the prediction. The graph shows that there is a large variation in predicted and actual values in the months from November to January. This variation might be because of number of holidays in this period of the year.

Comparison of the developed model with the second order regression model developed in a previous research on FM -158 Phase II highway project

A comparison of the developed first order linear regression model was compared with the second order linear regression model developed by Lee et.al in 2007. The second order model developed using the data collected from FM-158 Phase II project is as shown below –

$$NWD = 8.361 + 0.824\,(WDH) + 1.974\,(P) - 0.022\,(P * T)$$

Where, WDH = Weekend Days + Holidays – 8, P= Precipitation in inches, T = Temperature in degree Fahrenheit

As seen from the graph below, the model from the current research is at the lower side of the prediction which is more advantageous to the contractors as they will allocate more non-working days at the bidding stage. The difference in the predicted values might be because the two models were generated using different sample sizes of data. The following graph shows the comparison of the predicted values from the two models and the actual working days during the course of the project from April 2002 to July 2004.

The graph below also shows that the actual number of non-working days is very close to both the models, the model developed in this research is to the lower side of the prediction whereas, the second order regression model developed by Lee et.al in 2007 is slightly to the higher side of the prediction.

A combined regression analysis using the data from FM-158 Phase II project and SH-6 project would probably develop a model which is more close to the actual values.

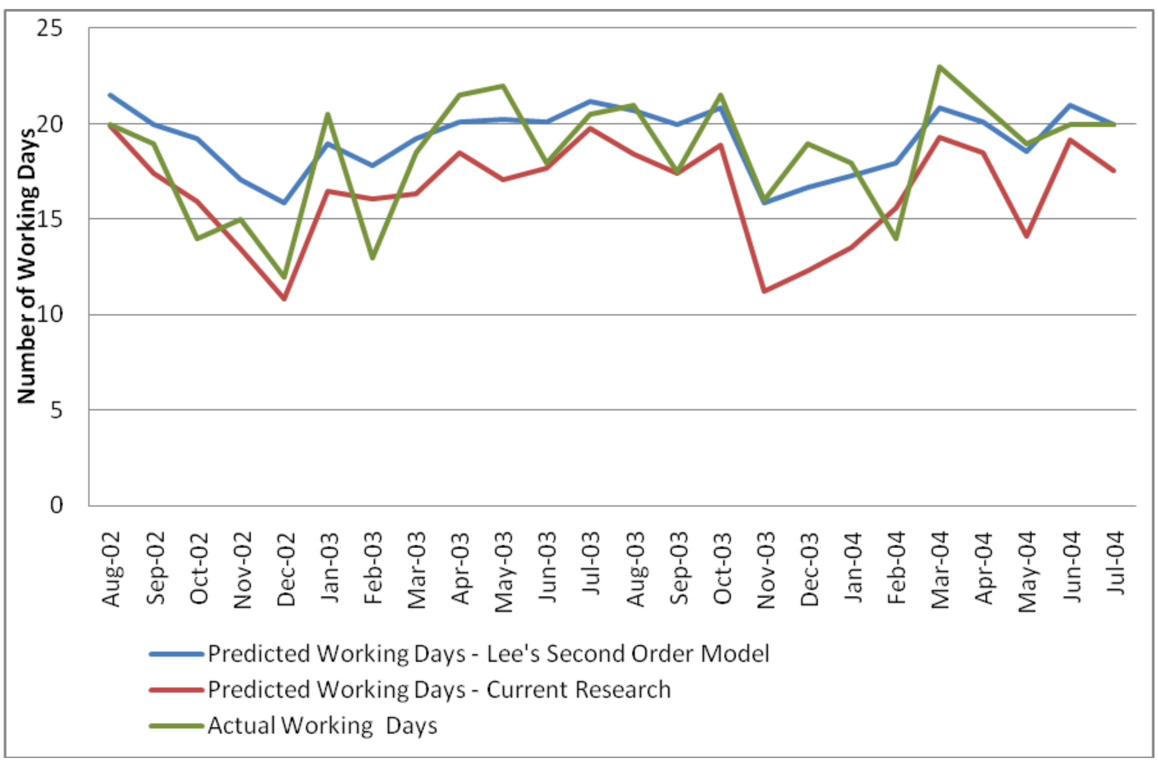

Figure 2: Comparison of Actual Values, Predicted values from Lee's second order Model and Current Model

Conclusions

The regression model developed in this research has an R squared value of 0.748 and an adjusted R square value of 0.719 shows that around 72% of the variability in the data has been accounted for in the model. Independent variables; precipitation, temperature and the calculated variable weekend days and holidays proved to be significant predictors of number of non-working days. Wind Speed, the least significant predictor was removed from the model by backward elimination procedure.

As an attempt to validate the model, the predicted values from the model using the historic weather data was compared to the actual number of working days which showed only 1.2% deviation from actual. The model was also applied to FM-168 Phase II project and compared to the actual values. The model showed 11% deviation to the lower side of prediction from the actual number of working days. The second order regression model developed by Lee et.al showed a 3.5% deviation from the actual values to the higher side of the prediction. A regression analysis using the combined data of FM-158 Phase II project and SH-6 project would possibly generate a model which is much closer to the actual values. The model developed in this research is simple, efficient and easy to use. All input data required in this process is easily available. The model is expected to serve engineers and contractors as a useful tool in predicting number of non-working days and hence the project duration of any future highway construction projects, provided the start date and duration of the project are known.

References

D&B. (2008, 10 19). *Highway and Street Construction Industry*. (D&B Company) Retrieved October 19, 2008, from Hoovers: http://premium.hoovers.com/subscribe/ind/fr/profile/basic.xhtml?ID=96

Kenner, S., Johnson, L. R., Matt, A. S., Miller, R. J., & Salmen, A. J. (1998). *Development of working day weather chart for transportation construction in South Dakota (Report No. SD97-07-F).* South Dakota Department of Transportation, Retrieved June 11, 2008, from http://ntl.bts.gov/lib/9000/9000/9083/SD97-07final.pdf

Lee, W. & Woods, P. K. (2008). *Predicting Non-Workdays Due toAdverse WeatherConditions for Highway Projects in Brazos County, Texas Using 2nd Order Linear Regression.* Unpublished master's professional paper, Department of Construction Science, Texas A&M University.

Longnecker, M. T., & Ott, L. R. (2001). *An Introduction to Statistical Methods and DataAnalysis* (5th ed.). Australia ; Pacific Grove, CA.: Duxbury.

Woods, P. K., Jue, W., Speed, M., & Burt, R. (2006). *Predicting Non-Workdays for Highway Projects in Brazos County, Texas from Historical Weather Records and Daily Project Work Reports*. Unpublished master's professional paper, Department of Construction Science, Texas A&M University.

Xi, Y., Balaji, R., & Molenaar, K. (2005). *Quantify construction delays due to weather.*Retrieved June 11, 2008, from http://forecast.colorado.edu/~apipatta/delay/