

Using Seasonal Forecast Data to Predict DOT Costs

Dan Pollock, Brad Temeyer, Lisa Strehlow, Mindy Albrecht

Hypothesis:

Seasonal forecast data can be used by the Iowa Department of Transportation (DOT) to estimate cost of maintaining roads in winter months so that over or under budgeting does not occur.

Procedure:

The Iowa DOT has the state divided into six regions labeled northwest, north central, northeast, southwest, south central, and southeast. Using a map of those regions, four to six observation sites distributed across the region were chosen as representation sites for the region. These sites selected were chosen for their location within the region and observational data available.

Observed data was used for each site from January- March 1998, and January - March 1999. For each month temperature, precipitation, snowfall, and Iowa DOT budget was used. The data for each site within a region was averaged to get a regional average for each month. The latitude and longitude for each site was also recorded and they too were averaged to find a latitude and longitude point to represent each region for use with the seasonal forecast data.

The monthly outputs from the model give data on temperature, precipitation, and 200 millibar winds on a grid with spacing of 2.8125 degrees of either latitude or longitude. The seasonal forecast program also runs hind casts of the same parameters over a twenty-year span (1979-1999) for six months following the initial conditions used

to drive the model. This was the information we used. Unfortunately, the precipitation was in units that caused the results to be close enough to zero that the program output was zero and that data was useless. Therefore, we only compared temperature with budget costs from the DOT. Six model grid points surrounded all the zone average points for our study. The monthly average temperature for each grid point was found for the six months mentioned earlier. The zone average temperature for those months was then found by using the four grid points closest to the zone point. Linear interpolation was used to first find the temperatures at the two lines of longitude on either side of the zone point at the same latitude as that zone point. Using the change in temperature with the change in longitude at that latitude, we were able to find the mean monthly temperature for that zone point. This process was repeated for all six-grid points.

It was assumed that all six regions were about equal area of responsibility and in size coverage area. Because information road mileage information for each region was not available for each region, it was assumed that each region had the same number of road miles.

Data Analysis:

The original intent of this study was to use a multiple linear regression to find a relationship between forecasted temperature and precipitation and the DOT costs. However, since the precipitation data could not be used, it was decided to first try statistically to show how useful the temperature forecast data is, and to then relate that temperature data to the DOT costs.

To find how useful the temperature forecast was, we compared it to our observed data. The residual, or the difference between the zone average forecasted temperature

and the actual zone average temperature was found for each month and used for each region of the study. It was determined that the best way to group the data was by lead-time of the forecast. Data for the month following the initial conditions were grouped in “One Month Lead Time.” Data for two months following initial conditions were grouped together, as was data for three months following initial conditions. The average difference between forecast and observed temperatures was found for each lead-time group. The standard deviation of difference was also found. Using this data, a 90% confidence interval was found for the difference between forecasted and observed temperature for each group in the hopes that zero be within the interval.

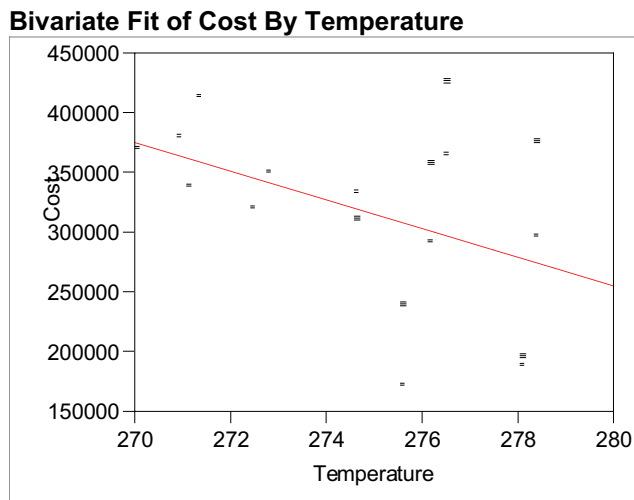
The same lead-time grouping was used for the linear regression. The 1998 data was used to create equations for predicting costs that could then be applied to 1999 data for testing. Data for temperature and cost were put into the put into the statistical program JUMP and an equation relating cost and temperature was found for each of the three lead-time groups. These equations were then used by plugging in 1999 forecasted temperature data and getting out a forecasted cost. We compared the residual between the forecasted costs and the actual costs. The average difference for each lead-time group was found and the confidence interval calculated as in the temperature portion of the study, again with hopes of finding zero in the confidence interval.

Results:

The temperature comparison for the one month lead time group showed an average difference of 2.965 degrees Kelvin with a 90% confidence interval of 2.67 to 3.26. This means that we can saw with 90% certainty that the forecast will overestimate the temperature by 2.67 to 3.26 degrees Kelvin. 2.965 degree Kelvin difference is

approximately a 5 degree Fahrenheit difference which is not bad for January, February, and March where temperatures can vary widely depending on snow cover and cloud conditions. The two-month results were not quite as good with an average difference of 5.669 Kelvin or around 10 degrees Fahrenheit. The confidence interval, 4.49 to 6.85, is wider because of the smaller sample size. The three-month forecast gets slightly worse with an average difference of 6.479 Kelvin or 11.66 degrees Fahrenheit. However, the standard deviation gets lower here than the previously mentioned two month interval so the confidence interval gets smaller going from 5.59 to 7.37 Kelvin degrees.

The one-month lead-time group for 1998 gave the equation relating cost and mean monthly temperature as follows:



— Linear Fit

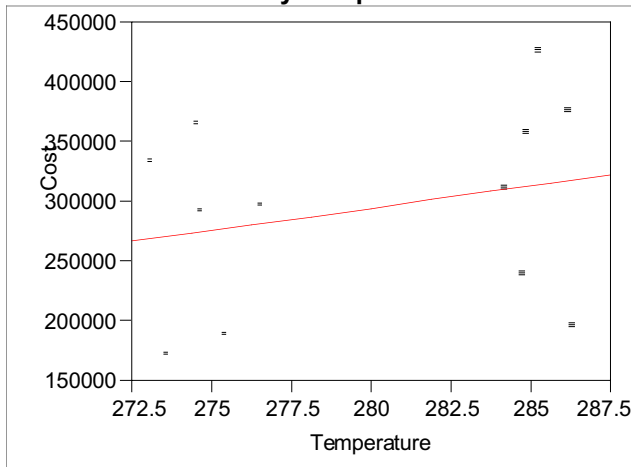
Linear Fit

Cost = 3632228.1 - 12060.943 Temperature

When this equation was applied to 1999 forecasted temperatures and the resulting cost compared to the actual cost, an average difference of minus \$59,992.71 was found. The 90% confidence interval was from -\$89,127.49 to -\$30,857.93.

The two-month group for 1998 showed a relationship of this:

Bivariate Fit of Cost By Temperature



— Linear Fit

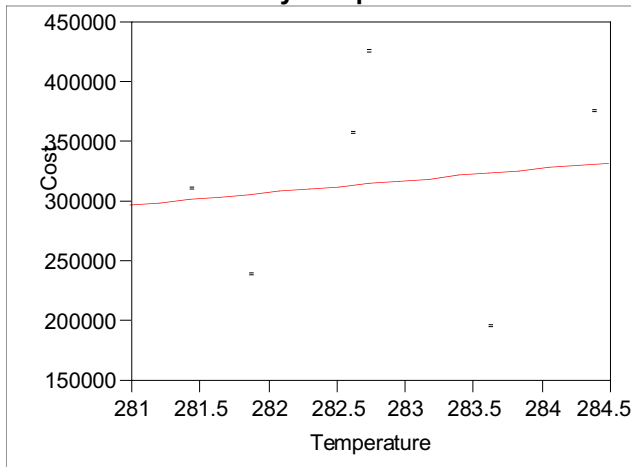
Linear Fit

$$\text{Cost} = -732422.9 + 3669.2974 \text{ Temperature}$$

This shows an actual increase in cost with an increase in temperature, which makes sense up to a point. Because warmer air can hold more moisture, one would expect snow removal costs to increase from colder temperatures to temperatures. However, when it is warmer than 32 degrees, one would expect the snow removal costs to be significantly less, which cannot be represented by a linear model. Therefore, it is important to keep in mind this only applies for temperatures below freezing (273 degrees K). The average difference between the forecasted cost and the actual cost was \$2,762.31 with a 90% confidence interval from about minus \$22,000 to around \$28,000. This was the best fit of the three.

The third test run yielded these results:

Bivariate Fit of Cost By Temperature



— Linear Fit

Linear Fit

$$\text{Cost} = -2.5531e6 + 10144.61 \text{ Temperature}$$

Again we see a slight increase in cost as temperature rises. This time the average distance was \$28,256.86 with a confidence interval from \$13,000 to almost \$43,000.

Conclusions:

When looking at the results of this study, we must conclude that no useable equation for forecasting cost was found for any amount of lead-time. When looking at the results for the temperature forecast, they seem good. This leads us to believe that there is not a direct relationship between temperature and cost. It would be interesting to see what the results would have turned up if useable precipitation data had been available from the seasonal forecast information. This data may be able to be used in future studies exploring the idea of using seasonal forecasts to estimate the future costs to be incurred by the Iowa DOT.