Statistical Modeling and Analysis on the Environmental Impact of Airport Deicing Activities

Huiyuan Fan[1], Prashant K. Tarun[2], Dachuan T. Shih[3], Seoung Bum Kim[4],
Victoria C.P. Chen[1], Jay M. Rosenberger[1], Dan Bergman[5]

[1]Department of Industrial and Manufacturing Systems Engineering
University of Texas at Arlington
Arlington, TX 76019-0017, USA

[2]Steven L. Craig School of Business
Missouri Western State University
St. Joseph, MO 64507, USA

[3]PFS, Tenet
13737 Noel Road, Suite 100
Dallas, Texas 75240, USA

[4]Department of Industrial Systems and Information Engineering
Korea University
5-ga Annam-dong, Sungbuk-gu
Seoul, Korea

[5]Dallas-Fort Worth International Airport
3200 East Airfield Drive
DFW Airport, TX 75261, USA

COSMOS Technical Report 09-01

Abstract

This paper presents an analysis on the environmental impact of airport deicing activities at Dallas-Fort Worth (D/FW) International Airport. Aircraft deicing uses a spray of aircraft deicing and anti-icing fluids (ADAF). ADAF has a high concentration of ethylene/propylene/diethylene glycol, which shears off airfoil surfaces during takeoff and drips to the runways during taxiing. A significant portion of the glycol runs off and mixes with the airport's receiving waters during heavy deicing periods causing an increase in bacterial growth and a subsequent reduction in dissolved oxygen (DO) in the receiving waters. Moreover, a high concentration of glycol in the receiving waters causes an increase in the chemical oxygen demand (COD), which is a measure of water quality. In this research, statistical methods for data mining were employed to evaluate the impact of airport deicing activities on DO and COD in the receiving waters immediately surrounding D/FW Airport. In particular, decision tree models were developed to determine important explanatory variables for predicting levels of DO and COD. The impact of glycol usage on DO was apparent as every decision tree had at least one group with a median DO below 4.0 mg/l. These low-DO groups were associated with high glycol usage. The decision tree modeling and analysis of COD determined *North-South wind*, *glycol usage* at a deicing pad, and *monitoring site* to be significant explanatory variables. These results are crucial to D/FW Airport in their goal to minimize the potential adverse impact of deicing activities on the water quality in waterways proximate to D/FW Airport.

Keywords: Airplane deicing and anti-icing; Water quality; Data mining; Decision trees; Glycol; Dissolved oxygen; Chemical oxygen demand

1. Introduction

Airplanes on the ground or in flight are susceptible to ice formation under various atmospheric and operational conditions, such as frost, snow, freezing precipitation, etc. (Revitt and Worrall, 2003; Revitt et al., 2001; Corsi et al., 2006; Switzenbaum et al., 1999; FAA Report, 1996; Leist et al., 1997). Ice that adheres to the surface of the airplane wing will hinder the smooth flow of air, thereby greatly degrading the ability of the wing to generate lift. Large pieces of ice that dislodge while the airplane is in motion can get caught in a turbine engine or may impact moving propellers with a potential to cause catastrophic failure. A thick layer of ice can also lock up the control surfaces impairing its functionality. Due to these potentially dangerous consequences, deicing and anti-icing are performed meticulously at airports during winter conditions.

The application of glycol-based aircraft deicing and anti-icing fluids (ADAF) has been the worldwide standard for airplane deicing/anti-icing at airports. These fluids contain ethylene/ propylene/diethylene glycol, water, and proprietary additives (the additive packages). Fluid that drips onto the ground or shears off the airplane during take-off can runoff to receiving surface waters or into the groundwater system. This has the potential to cause adverse environmental impacts in the form of an increased aquatic toxicity and oxygen depletion in the receiving waters. Aquatic toxicity is potentially caused by the additive packages in ADAF, and oxygen depletion occurs due to a high biochemical oxygen demand (BOD) and chemical oxygen demand (COD) of glycols in ADAF (Corsi et al., 2006; U.S. Environmental Protection Agency, 2000). BOD and COD are measures of water quality that quantify the amount of oxygen consumed in the biochemical and chemical oxidation processes, respectively (Masters, 1997). Thus, an increase in BOD and/or COD levels in water results in a decrease in the dissolved oxygen (DO) level due to consumption of oxygen in the oxidation process caused by the biochemical and chemical agents (Corsi et al., 2006).

Dallas/Fort Worth (D/FW) International Airport, located in north-central Texas, USA, is one of the world's largest and busiest international airports (Corsi et al., 2006). Typically, D/FW Airport witnesses sporadic deicing periods every winter season requiring airplanes to be deiced/anti-iced in compliance with Federal Aviation Administration (FAA) safety regulations. Deicing activities at D/FW Airport have received much attention in recent years, especially after an ecological mishap in 1999 when deicing led to a significant amount of glycol runoff into Trigg Lake (a local irrigation reservoir that receives D/FW Airport runoff), resulting in a fish kill in the lake.

In the wake of this mishap, D/FW Airport upgraded its ADAF collection facilities by constructing eight deicing source isolation pads at which spent ADAF runoff is channeled into the airport's reverse osmosis wastewater treatment system. This is hypothesized to capture about 80% of the spent ADAF runoff. The remaining 20% is due to "drip and shear" as airplanes taxi to the runway and take-off. Spent ADAF runoff due to drip and shear may discharge into local receiving waters, which can lead to a detrimental effect on the water quality and aquatic life. To compensate for environmental impact of "drip and shear" in Trigg Lake, D/FW Airport installed 17 aerators in the lake to maintain proper DO levels and avoid any recurrence of the 1999 environmental mishap.

To monitor DO levels, D/FW Airport, in collaboration with the United States Geological Survey (USGS), implemented the collection of water quality data at nine sites in waterways surrounding the airport: a rural reference site at Blessing Branch (BLSN); an urban reference site on Big Bear Creek at Euless/Grapevine road near Grapevine, TX (REF); an airport site at Outfall #19 on an unnamed tributary off Big Bear Creek near Euless, TX (OF19); an airport site draining into Trigg Lake (IN); three sites within Trigg Lake (S1, S2, S3); an airport site downstream from Trigg Lake (OUT); and a downstream site on Big Bear Creek at SH 183 near Euless, TX (DNST). BLSN and REF are reference sites because they are upstream from the airport and consequently are not affected by airport activities. However, water quality at REF may be subject to activities in the surrounding urban area, while BLSN should represent natural water quality. DO levels at sites S1, S2, and S3 within Trigg Lake, and sites OUT and DNST downstream from Trigg Lake are impacted by the aerators. By contrast, sites IN and OF19 remain subject to airport activities without any remediation. Figure 1 provides a schematic diagram of the locations of the USGS monitoring sites (dark circles) relative to the airport. The eight deicing pad locations (squares) are also shown. Each pad location has multiple slots on which to deice airplanes.

Figure 1 here

With these data, D/FW Airport monitors its "end-of-pipe" DO, BOD, and COD levels to ensure that the airport's deicing/anti-icing practices are environmentally-friendly and state-of-the-art. The key to achieving the best deicing/anti-icing practice is to understand the interrelationships among deicing, water quality, meteorological, and several other relevant variables. This research works toward identifying explanatory variables that affect DO and COD levels in the airport's receiving waters significantly, and thus assisting D/FW Airport management to improve various aspects of the current practice. One of the overarching goals of this paper is to provide D/FW Airport with the tools to improve its current deicing/anti-icing practices and explore new deicing/anti-icing chemicals, technologies and methodologies with a potential to minimize water quality risk.

2. Materials and Methods

2.1 Data

The following data sets were employed for this study. The first two sets of data were collected by D/FW Airport in collaboration with the USGS. The third set was collected by the airport, and the fourth one was modified for the analysis.

1. *USGS Continuous Monitoring for Six Sites*: Table 1 shows the variables that were monitored at sites BLSN, REF, DNST, OF19, IN, and OUT along with the dates.

Table 1 here

2. *USGS Manual Sampling*: The COD data were collected at sites REF, DNST, OF19, IN, and OUT, during deicing events. The major deicing periods considered were: Jan 2003: 01/12/03–01/13/03, Feb 2003: 02/26/03–02/28/03, and Feb 2004: 02/14/04–02/16/04. The

minor deicing period considered was Dec 2002: 12/30/02–12/31/02, and the test deicing event was Aug 2003: 08/26/03.

3. *Airport Deicing Activities*: The daily ethylene and propylene glycol usage, and deicing pad usage at 8 deicing pad locations were recorded. Deicing pad locations and time durations are shown in Table 2.

Table 2 here

4. *Airport Meteorology*: The hourly data for temperature, precipitation, wind speed & direction were taken from TDL U.S. and Canada Surface Hourly Observations (http://dss.ucar.edu/datasets/ds472.0/) for the DFW stations located at latitude from 32.5N to 33.5N and longitude from 96.5W to 97.5W for the interval 2002 - 2004.

The data for USGS Continuous Monitoring were automatically recorded via sensors approximately every 15-20 minutes. The sampling times for USGS Manual Sampling varied from several minutes to a few hours. Glycol usage from airport deicing activities was aggregated by airline and deicing pad location for each day. Meteorological data were recorded hourly. The following three data sets were constructed to conduct the analyses.

1. Data Set 1**:** Hourly-averaged USGS continuous monitoring data for the six sites.
2. Data Set 2: Merged DO data from USGS continuous sampling, airport deicing activities, and airport meteorology for days on which deicing occurred during the 2002 – 2003 and 2003 – 2004 deicing seasons.
3. Data Set 3: Merged COD data from USGS manual sampling, airport deicing activities, and airport meteorology for deicing event days in December 2002, January and February 2003, and February 2004.

Since deicing activities were only recorded by day for Data Sets 2 and 3, observations taken multiple times per day needed to be combined to provide a daily measure. Specifically, the daily minimums of hourly-averaged DO were calculated, the daily maximums of COD were calculated, and daily averages were calculated for all other variables.

For the meteorological data, the wind speed and wind direction variables have been transformed as follows, for easier modeling. Given wind observations recorded as wind speed u_i and wind direction θ_i, the mean east-west component V_e and north-south component V_n of the wind are computed by the following equations:

$$V_e = -\sum u_i \sin(\theta_i),$$

$$V_n = -\sum u_i \cos(\theta_i).$$

Positive values of V_e indicate a wind component blowing from east to west, while negative values indicate a wind component blowing from west to east. Similarly, positive values of V_n indicate a wind component blowing from north to south, while negative values indicate a wind component blowing from south to north.

2.2 Decision Tree Method

A decision tree method is used to analyze the adverse environmental impacts of deicing and anti-icing activities at D/FW Airport. The decision tree method was first developed by Breiman et al. (1984) for both classification and regression, and has now become an extremely popular data mining tool (Huo et al., 2006). Decision tree models are flexible in that the models can efficiently handle both continuous and categorical variables in the model construction. From the algorithmic point of view, the decision tree method has a forward stepwise procedure that adds model terms and a backward procedure for pruning and conduct variable selection by only including useful variable in the model (Breiman et al., 1984). The output of decision tree models is a hierarchical structure that consists of a series of if-then rules to predict the outcome of the response variable, thus facilitating the interpretation of the final model. Figure 2 shows an example of the regression tree model.

Figure 2 here

At each intermediate node (ovals in Figure 2) of the tree, a question is asked about the state of the variables (e.g., X1, X2, and X3) of input data. The data set that satisfies the question goes left in the branching and right if it fails to meet the criterion. Based on the values (e.g., C1, C2, and C3) of the variables, every data point ends up in one of the terminal nodes (rectangles in Figure 2) of the tree. We can then determine the criteria for each terminal node by backtracking up the tree to the top node. For example, in Figure 2, the first terminal node (the terminal node farthest to the left) backtracks up the left edge of the tree, yielding the following rule: "If X1 is greater than C1 and X2 is greater than C2, then the average (or median) values of the response variable is Z1." Other terminal nodes in the tree can be interpreted similarly.

3. Results and Discussions

3.1 Analysis of DO at Monitoring Sites

Our first statistical analyses were performed on the 2004 data from Data Set 1 to study DO alone for the purposes of identifying those variables that are related to DO at the different sites. Given that the continuously-sampled data were observed over time, it was important to address potential serial correlation between observations over time. To do this, an autocorrelation analysis was conducted for DO. The autocorrelation analysis led to the inclusion of time-lagged variables in the set of explanatory variables. Further, the decision tree analyses were conducted at the six monitoring sites- BLSN, REF, DNST, OF19, IN, and OUT. The results of autocorrelation and regression tree analyses on DO are presented next.

3.1.1 Autocorrelation Analysis of DO

Since the amounts of DO were observed over time, its autocorrelation was analyzed first. Lag plots were generated to show the serial correlation in the DO measurements at the six monitoring sites listed above. To gain some clarity on the concept of time lags, if DO at time $t-1$ affects DO at time t, then there is an autocorrelation with a single time lag. Similarly, if DO at times $t-1$ and $t-2$ affects DO at time t, then there is an autocorrelation with two time lags. The autocorrelation analysis seeks to identify the number of time lags required for an adequate representation of the serial correlation in the DO measurements. Figure 3 shows the correlation between DO measurements and DO with one through six time lags at site OF19, for example. The results for all six monitoring sites showed that only the first few lags were important. Furthermore, a partial autocorrelation analysis, which seeks to measure the degree of correlation between neighboring data observations in a time series, was also done for DO for each site.

Figure 3 here

Figure 4 illustrates partial autocorrelation of DO at site OF19, where longer lines indicate higher correlations. The longest line is for time lag 1, and the next longest is for time lag 2. Higher time lags have similar yet lower correlations. The partial autocorrelation analysis for all six monitoring sites indicated a necessity to include two time lags.

Figure 4 here

3.1.2 Regression Tree Analysis on DO

Regression tree models were developed for DO at the six monitoring sites using the Classification and Regression Trees (CART) software (www.salfordsystems.com). Given the results of the autocorrelation analysis it was decided to include all variables with two time lags. Consequently, regression trees were constructed to model DO as a function of water temperature, discharge rate, and precipitation each with two time lags as well as two time lags of DO. The regression tree topologies for all six monitoring sites are summarized in Table 3.

Table 3 here

For variable selection, CART provides “variable importance scores.” The variable receiving a score of 100 is considered to be the most influential variable for prediction, followed by other variables in the order of their variable importance scores. The ranking of important explanatory variables to predict DO at six monitoring sites of D/FW Airport is presented in Table 4. Across the sites the time-lagged DO variables are the most important followed by the water temperature variables. On the other hand, the discharge rate and precipitation variables have much lower importance. It is interesting to note that the water temperature variables are ranked in ‘reverse’ time order at REF, OF19, IN and OUT with the lagged time variables being more important than the non-lagged time variables. This indicates a time-lagged effect of water temperature on DO. In other words, DO in the current hour is influenced by the water temperature in the previous two hours.

Table 4 here

A detailed analysis was performed for the water temperature variable because it was identified as an important explanatory variable. Tables 5 through 10 provide splits involving water temperature variables in the resulting regression trees for the six monitoring sites. They are categorized into five groups: <10°C, 10-15°C, 15-20°C, 20-25°C, and >25°C. These splits indicate temperature values around which larger changes in DO are observed.

Since BLSN is the rural reference site, its pattern of water temperature splits could be interpreted as the “natural” pattern. It was immediately observed that REF, the urban reference site, had a very different pattern, with much higher temperature splits. Among the sites potentially impacted by airport deicing activities, DNST, OF19, and OUT had water temperature splits similar in magnitude. The site IN exhibited a pattern that could be interpreted as a combination of the patterns at the urban reference site REF and the other downstream sites DNST, OF19, and OUT. Apparently, one possible impact of airport deicing activities was its significant influence on DO at slightly lower water temperatures as opposed to relatively higher water temperatures (> 18°C). Also, it was interesting to note that the airport monitoring sites

DNST, OF19, and OUT had temperature patterns that were more similar to the “natural” pattern of BLSN than the urban reference site REF.

Tables 5-10 here

3.2 Analysis of DO by Deicing Pad Locations

Of particular interest to the D/FW airport is the nature of association between DO/COD and glycol usage at eight deicing pad locations: taxiways EKS, WK, HY, Z, and C; and hold pads SE, SW, and NE. Knowledge of these relationships will enable D/FW Airport to manage the deicing activities so as to minimize its potential adverse effects on the quality of receiving waters surrounding the airport. This section extends the previous DO analyses to incorporate the effect of deicing activities, represented by glycol usage (total amounts of ethylene and propylene glycols), from Data Set 2. The glycol usage with a lag of one day was added as a potential explanatory variable as a result of an autocorrelation analysis.

Initially, a regression tree model was attempted for all eight deicing pad locations, but strong correlations of glycol usage across the deicing pads led to a model with one deicing location while listing some other deicing pad locations as competitive “surrogates.” The high correlation could be attributed to a simultaneous rise and fall in glycol usage across the deicing pad locations. It must be recognized that high correlations of glycol usage across the deicing pad locations do not imply that the average amounts of glycol are similar at all deicing pad locations; some locations conduct much more deicing than others. Intuitively, the high correlations are reasonable since deicing events occur based on the weather, so glycol usage at the deicing pad locations will go up and down based on the safety need for deicing. As a consequence of the high correlations, it was decided to study the impact of each deicing location separately.

Prior to the main analysis for DO, some preliminary analyses were conducted that led to the following interesting results: (a) the monitoring sites (BLSN, REF, DNST, OF19, IN, OUT), the total glycol usage, total glycol usage with one day lag, and meteorological variables (East-West wind component (EW) were important in predicting DO; and (b) the monitoring sites OF 19 and IN were influenced significantly by the deicing events. Interestingly, the result (b) was in line with how monitoring sites and deicing pads are actually located around the airport. The monitoring sites BLSN and REF are reference sites that are not impacted by airport deicing activities; while the DO levels at sites OUT and DNST are maintained by the aerators in Trigg Lake. As a result of those analyses, we chose to focus on sites OF19 and IN for investigating the adverse impacts of glycol usage on DO in the receiving waters of the D/FW Airport (i.e., higher glycol usage causes a decrease in DO). We conducted the main DO analysis by constructing the decision trees for each of the eight deicing pad locations. Least absolute deviation was used as the splitting criteria for regression trees because this is less sensitive to outliers than using least squares. The testing method used was 10-fold cross-validation. The response variable was DO, and the predictor variables were monitoring sites (OF19 and IN), total glycol, total glycol with one day lag, and meteorological variables East-West wind component (EW) and North-South wind component (NS).

Figure 5 shows the explanatory variables and structure of the resulting regression trees. Table 11 shows the ranking of important variables for each deicing pad location. The key findings are:

- Splitting information on glycol usage (or glycol usage lagged by one day) appears in the tree models for Taxiways WK, Z, and C, and SE Hold Pad. More importantly, extremely high glycol usage is seen to decreas*e* DO at monitoring site IN. However, moderately higher glycol usage at Taxiway C appears to increase DO at monitoring site OF19.

- A stronger north to south wind component contributes to lower DO, while a low east-west wind component may contribute to lower DO.

Figure 5 here
Table 11 here

Tables 12-15 exhibit the groupings of DO observations in the decision trees. Every decision tree includes at least one group with a median DO below 4.0, and all of these low DO groups are associated with high glycol usage. Although there are relatively few observations in these low DO groups, these are exactly the critical cases that D/FW Airport would like to avoid in order to minimize the adverse impact of deicing activities on water quality in the airport's receiving waterways.

Tables 12-15 here

3.3 Analysis of COD by Deicing Location

The relationship between COD at monitoring sites and combined total amounts of glycol usage was explored based on COD data in Data Set 3. Similar to the analysis in Section 3.2, the impact of glycol usage was studied separately for each deicing pad location. To address the potential time-lagged effect of deicing activities on COD, total amounts of glycol usage with a time lag of one day through six days were considered.

Regression trees were built for each deicing pad location as shown in Figure 6. Differences in the regression tree models could be seen in terms of the impact of deicing activities at specific deicing pad locations. Taxiways WK, HY, and C, and SE Hold Pad are important (tied with NS) in their decision trees, while Taxiways EKS and Z, and Hold Pads SW and NE hold little importance in their trees. Of the important deicing pad locations, Taxiways WK, HY, and C are all located in the northwest corner of the airport as can be seen in Figure 1. The SE Hold Pad location is to the south, on the east side of the airport, exactly opposite to the other three. Furthermore, the decision tree results indicate that the COD level significantly varies with each monitoring site, and sites IN and OF19 are affected significantly by deicing activities.

Figure 6 here

Table 16 exhibits the groupings of COD observations in the regression tree at Taxiway WK. Taxiways HY and C, and SE Hold Pad had mathematically identical groupings for COD as Taxiway WK. Although the majority of observations were in the first group with an acceptable COD (below 120) at monitoring sites REF, DNST, and OUT, there were a number of observations in other groups with significantly higher COD at sites IN and OF19.

Table 16 here

Table 17 shows the ranking of important variables in predicting COD. The results presented in this table demonstrate that the north-south wind component (NS) was the most important variable for the COD models across the eight deicing pad locations. In particular, a higher north to south wind speed yielded higher COD levels at monitoring sites OF19 and IN. On the other hand, monitoring sites, REF, DNST, and OUT were not affected at all by the NS wind. The east-west wind component (EW) was not as important in predicting COD levels. It was also found that other meteorological variables and time-lagged variables were not as important in predicting COD levels.

Table 17 here

4. Conclusions

This paper presented analyses of the potential adverse impacts of airport deicing/anti-icing activities on the water quality in the waterways surrounding D/FW Airport. A series of decision tree models were developed for studying DO in the receiving waters of D/FW Airport. The tree analysis on DO at six monitoring sites identified the time-lagged DO measurements as having the strongest relationship with DO, followed by the water temperature variables. The discharge rate and precipitation variables were not as important across all monitoring sites. The tree models constructed for predicting DO levels in the receiving waters of D/FW Airport due to deicing activities at eight deicing pad locations led to some interesting results. The impact of glycol usage on DO could be seen by the fact that every decision tree included at least one group with a median DO below 4.0 mg/l, and all of these low DO groups were associated with high glycol usage. These are exactly the critical cases that D/FW Airport would like to avoid in order to mitigate the impact of glycol usage on the water quality in the airport's receiving waters. Another interesting result was the impact of wind speed and direction on the DO level. A stronger north to south wind resulted in a lower DO level. This made physical sense given that the impacted waterways are south of the airport. The tree models used to identify patterns relating COD to total glycol usage in the airport deicing activities and several other meteorological variables across the eight deicing pad locations found that the amount of glycol usage, monitoring site, and again the north-south wind were important in predicting COD levels in the airport's waterways. Our future work includes the development of a data-driven optimization tool for deicing activities at D/FW International Airport.

Acknowledgements

We are grateful to Mr. Richard Reeter at D/FW International Airport for his useful comments. This study was supported by grants from D/FW International Airport.

References

Revitt, D.M., Worrall, P., 2003. Low Temperature Biodegradation of Airport De-icing Fluids. Water Science and Technology 48, 103-111.

Revitt, D.M., Worrall, P., Brewer, D., 2001. The Integration of Constructed Wetlands into a Treatment System for Airport Runoff. Water Science and Technology 44, 469-476.

Corsi, S.R., Harwell, G.R., Geis, S.W., Bergman, D., 2006. Impacts of Aircraft Deicer and Anti-icer Runoff on Receiving Waters from Dallas/Fort Worth International Airport, Texas, USA. Environmental Toxicology and Chemistry 25, 2890-2900.

Switzenbaum, M.S., Veltman, S., Schoenberg, T., Durand, G.M., Mericas, D., Wagoner, B., 1999. Workshop: Best Management Practices for Airport Deicing Stormwater. Publication No. 173, Amherst Water Resources Research Center, University of

Massachusetts, Massachusetts. (http://www.umass.edu/tei/wrrc/WRRC2004/pdf/Switz173.pdf).

FAA Report, 1996. Report on FAA Deicing Program at Guardia and O'Hare Airport. Report Number E5-FA-7-002, Office of Inspector General, Department of Transportation, Federal Aviation Regulation. (http://ntl.bts.gov/lib/1000/1500/1543/e5-fa-7-002.pdf)

Leist, D., Flowers, K., Bartley, G., Beebe, B., Giles, T., Ratcliff, E., 1997. Impacts of Deicing Fluids on Elijahs and Gunpowder Creeks Boone County, Kentucky.Kentucky Department For Environmental Protection, Division of Water, Frankfort, Kentucky. (http://www.epa.gov/owow/tmdl/examples/organics/ky_elijahgunpowder.pdf).

U.S. Environmental Protection Agency, 2000. Preliminary Data Summary Airport Deicing Operation. EPA 821-R-00-001, Office of Water, Washington, DC.

Masters, G. M., 1997. Introduction to Environmental Engineering and Science, 2nd Edition. Prentice Hall, Upper Saddler River, New Jersey.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Belmont, Wadsworth, California.

Huo, X., Kim, S.B, Tsui, K.L., Wang, S., 2006. FBP: A frontier-based tree-pruning Algorithm. INFORMS Journal on Computing 18, 494-505.

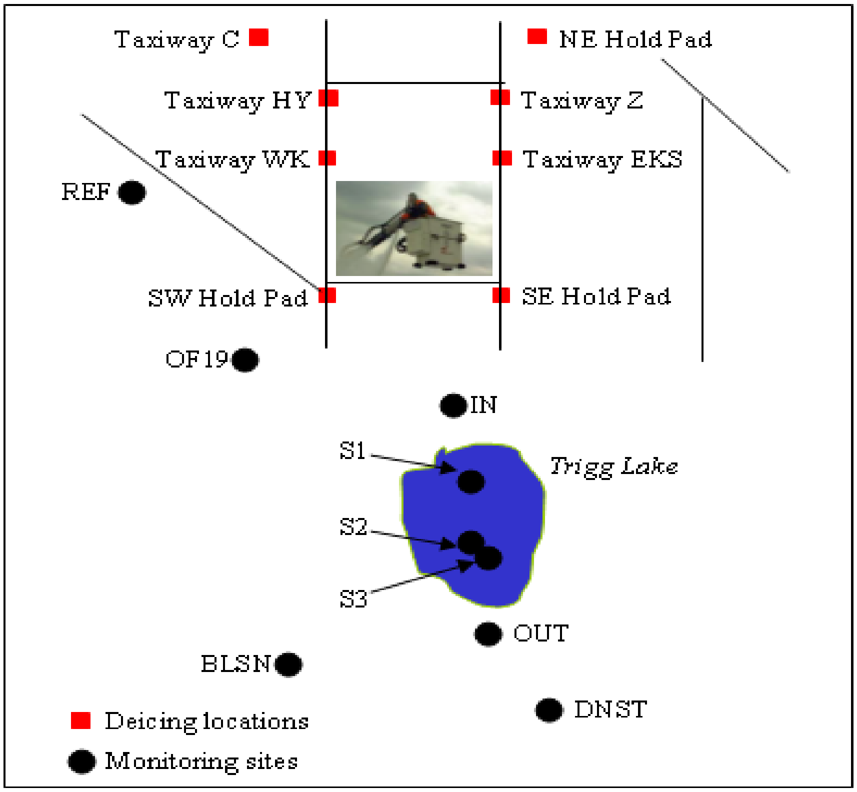


Figure 1: Schematic diagram of monitoring sites and deicing pad locations at D/FW airport. Circles represent the locations of the USGS monitoring sites and squares represent the deicing pad locations in D/FW Airport.

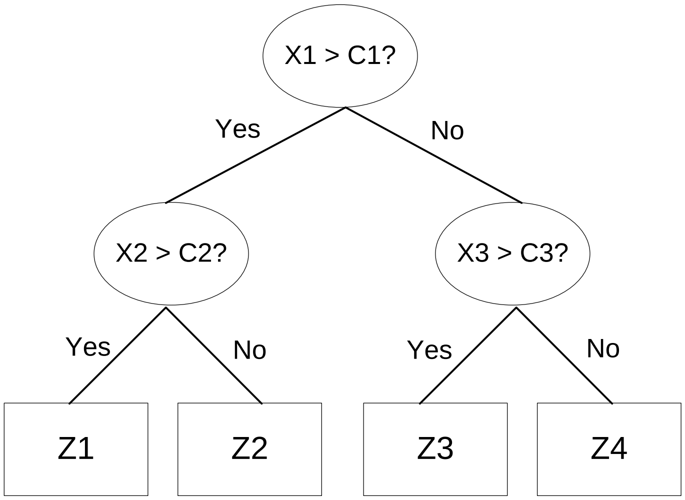


Figure 2: Example of the regression tree model. Oval nodes are the intermediate nodes and rectangles are terminal nodes. C1, C2, and C3 are the splitting values of the variables X1, X2, and X3. Z1, Z2, Z3, and Z4 are the average (or median) values of the response variable in the terminal nodes.

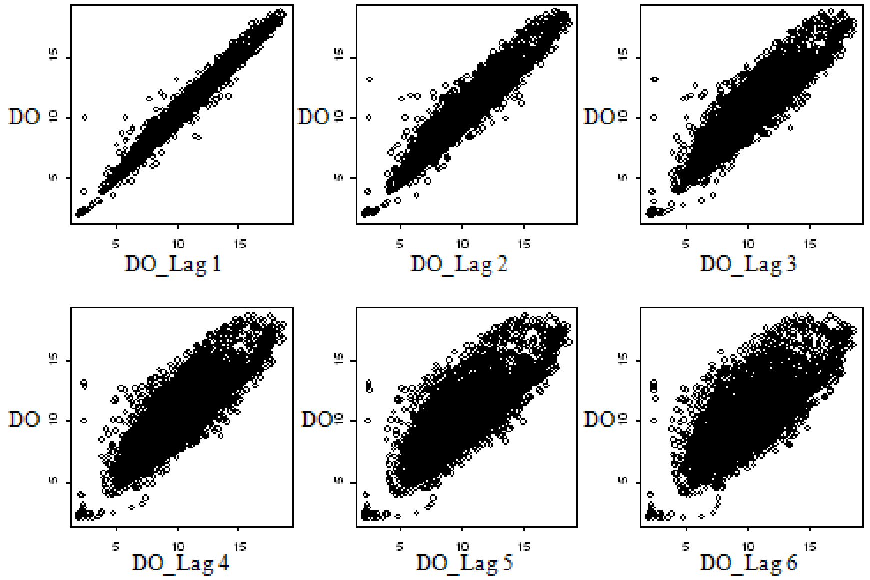


Figure 3: Scatter plots of pairs of DO values in the OF19 site with lags one through six.

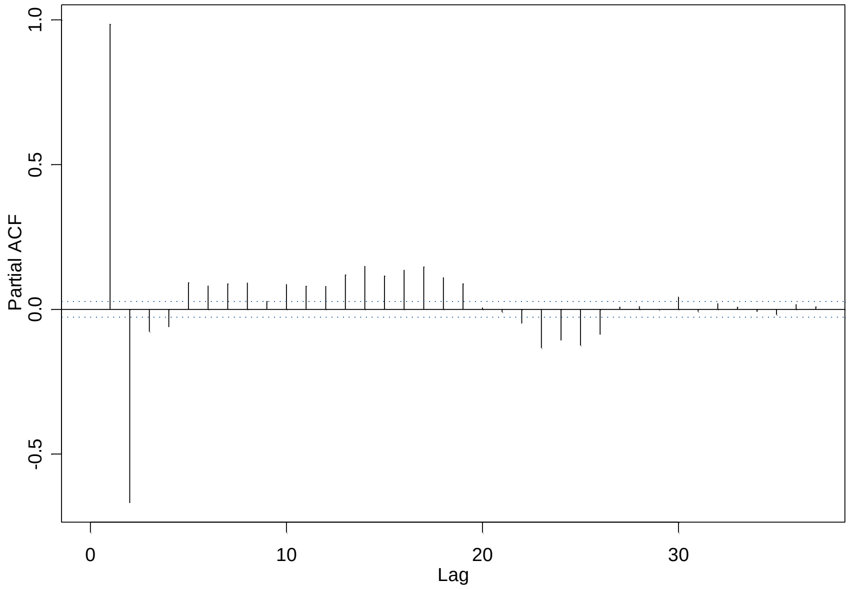


Figure 4: Partial autocorrelation plot of DO in OF19.

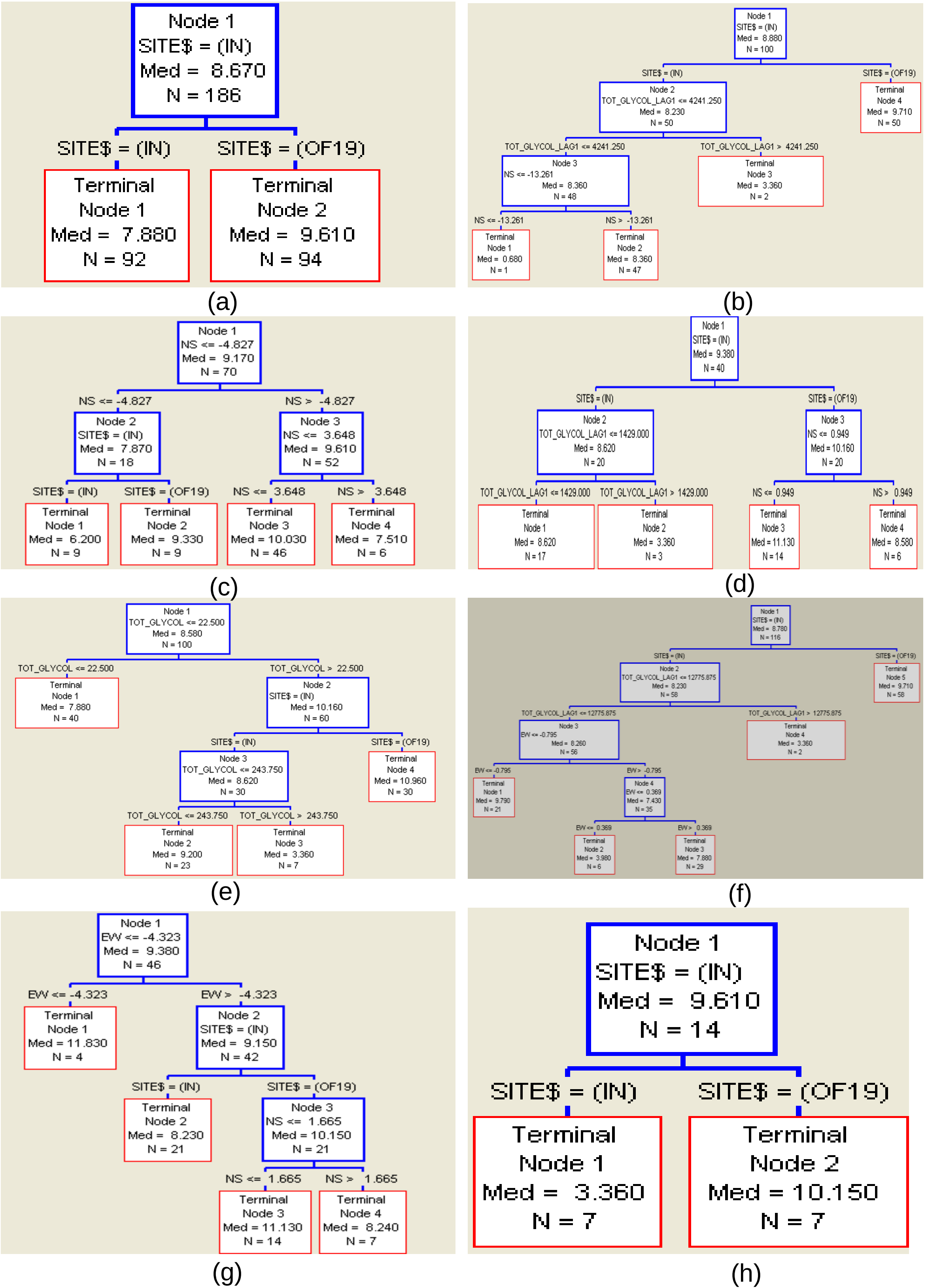
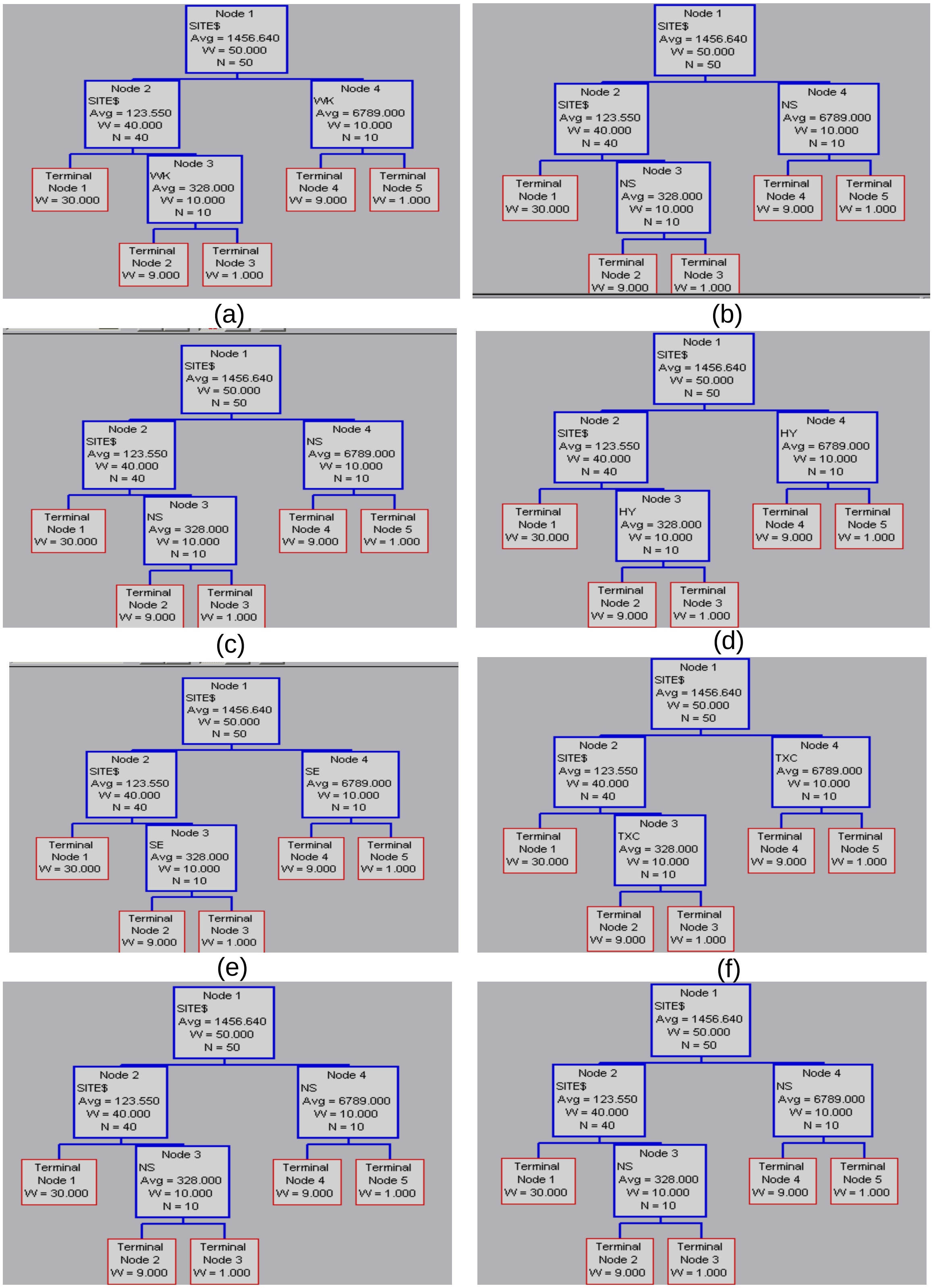


Figure 5: Tree structures from main DO analysis for deicing pad locations: Taxiways- (a) EKS, (b) WK, (c) HY, (d) Z, (e) C; Hold Pads- (f) SE, (g) SW, and (h) NE.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 6: COD tree structure for deicing pad locations: Taxiways- (a) EKS, (b) WK, (c) HY, (d) Z, (e) C, Hold Pads- (f) SE, (g) SW, and (h) NE.

Table 1: USGS continuous monitoring at six monitoring sites

Monitoring Sites	Variables Monitored	Duration
BLSN: (the rural reference site at Blessing Branch)	DO, water temperature, precipitation, and stage	10/22/03 - 04/15/04
REF: (the urban reference site on Big Bear Creek, Euless/Grapevine road, Grapevine, TX)	DO, discharge, water temperature, and precipitation	10/21/02–09/30/03, 10/01/03–04/16/04
DNST: (the downstream site on Big Bear Creek at SH 183, Euless, TX)	DO, discharge, and water temperature	10/21/02–09/30/02, 10/01/03–04/16/04
OF19: (at Outfall #19 on unnamed tributary off Big Bear Creek, Euless, TX)	DO, discharge, water temperature, and precipitation	10/21/02–09/30/03, 10/01/03–04/16/04
IN: (the site that drains into Trigg Lake)	DO, discharge, water temperature, and precipitation	10/21/02–09/30/03, 10/01/03–04/16/04
OUT: (the downstream site from Trigg Lake)	DO, discharge, water temperature, and precipitation	10/30/02–09/30/03, 10/01/03–04/16/04

Table 2: Deicing pad locations and time durations

Pad Locations	Duration
Taxiway EKS	10/27/2002 – 04/09/2003, 11/07/2003 – 03/12/2004
Taxiway WK	11/05/2002 – 04/09/2003, 12/11/2003 – 02/27/2004
Taxiway HY	11/28/2002 – 04/09/2003, 11/09/2003 – 02/26/2004
Taxiway Z	11/28/2002 – 02/26/2003, 12/14/2003 – 02/15/2004
Taxiway C	10/24/2002 – 02/28/2003, 12/01/2003 – 03/29/2004
SE Hold Pad	11/28/2002 – 03/25/2003, 12/08/2003 – 02/26/2004
SW Hold Pad	12/24/2002 – 02/27/2003, 12/24/2003 – 02/26/2004
NE Hold Pad	01/12/2003 – 02/26/2003, 02/14/2004

Table 3: Summary of regression tree topologies for monitoring sites

Site	Number of Terminal Node	Relative Error
BLSN	98	0.025
REF	99	0.014
DNST	110	0.009
OF19	95	0.022
IN	113	0.026
OUT	103	0.007

Table 4: Importance of variables to predict DO at six monitoring sites of D/FW Airport

Ranking	BLSN	REF	DNST	OF19	IN	OUT
1	DO (lag1)	DO (lag1)	DO (lag1)	DO (lag1)	DO (lag1)	DO (lag1)
2	DO (lag2)	DO (lag2)	DO (lag2)	DO (lag2)	DO (lag2)	Disch (lag1)
3	Temperature	Temp (lag2)	Temperature	Temp (lag2)	Temp (lag1)	Discharge
4	Temp (lag1)	Temp (lag1)	Temp (lag2)	Temp (lag1)	Temperature	Temp (lag1)
5	Temp (lag2)	Temperature	Temp (lag1)	Temperature	Temp (lag2)	Temperature
6	Precipitation	Disch (lag2)	Disch (lag2)	Disch (lag2)	Disch (lag2)	Precip (lag1)
7	Precip (lag1)	Discharge	Disch (lag1)	Disch (lag1)	Discharge	Precipitation
8		Disch (lag1)	Discharge	Precip(lag1)	Disch (lag1)	
9		Precip(lag2)		Precip(lag2)	Precip(lag1)	
10		Precip(lag1)		Precipitation	Precip(lag2)	
11		Precipitation			Precipitation	

Table 5: Splits involving temperature-related variables at BLSN

	<10°C	10-15°C	15-20°C	20-25°C	>25°C
Temperature		12.7, 12.9	15.3, 18.1, 18.6, 19.1	20.3, 21.3	
Temp (lag1)					
Temp (lag2)			17.6		

Table 6: Splits involving temperature-related variables at REF

	<10°C	10-15°C	15-20°C	20-25°C	>25°C
Temperature			18.2, 20.0	20.8, 22.0	25.8
Temp (lag1)			18.4		
Temp (lag2)			17.8	20.2	

Table 7: Splits involving temperature-related variables at DNST

	<10°C	10-15°C	15-20°C	20-25°C	>25°C
Temperature		10.6, 11.0, 11.1,11,4	17.9	22.9	
Temp (lag1)		14.8			
Temp (lag2)		10.9, 11.5	17.5		

Table 8: Splits involving temperature-related variables at OF19

	<10°C	10-15°C	15-20°C	20-25°C	>25°C
Temperature		10.2	18.4		
Temp. (lag1)					
Temp. (lag2)		10.9	16.1, 19.1		

Table 9: Splits involving temperature-related variables at IN

	<10°C	10-15°C	15-20°C	20-25°C	>25°C
Temperature	5.3	14.3		21.3, 24.2, 25.0	26.0, 26.6
Temp (lag1)				20.4	
Temp (lag2)			16.2		

Table 10: Splits involving temperature-related variables at OUT

	<10°C	10-15°C	15-20°C	20-25°C	>25°C
Temperature		10.0, 14.6			
Temp (lag1)					
Temp (lag2)		14.2	18.0		

Table 11: Main DO Analysis: Summary of important variables to predict DO

Ranking	Taxiway EKS	Taxiway WK	Taxiway HY	Taxiway Z	Taxiway C	SE Hold Pad	SW Hold Pad	NE Hold Pad
1	Site	Site	NS	NS	Glycol	EW	NS	Site
2		Glycol lag	Site	EW	Site	NS	EW	
3		NS	EW	Glycol lag	Glycol lag	Site	Site	
4		EW		Site		Glycol lag	Glycol	
5		Glycol		Glycol		Glycol		

Table 12: Groupings of DO observations in the Taxiway WK decision tree model.

Group	Site	WK Glycol Lag	NS	Median DO	# Observations
1	IN	≤ 4241.25	≤ –13.261	0.68	1
2	IN	≤ 4241.25	> –13.261	8.36	47
3	IN	> 4241.25		3.36	2
4	OF19			9.71	50

Table 13: Groupings of DO observations in the Taxiway Z decision tree model.

Group	Site	TxZ Glycol Lag	NS	Median DO	# Observations
1	IN	≤ 1429		8.62	17
2	IN	> 1429		3.36	3
3	OF19		≤ 0.949	11.13	14
4	OF19		> 0.949	8.58	6

Table 14: Groupings of DO observations in the Taxiway C decision tree model.

Group	Site	TxC Glycol	Median DO	# Observations
1		≤ 22.5	7.88	40
2	IN	(22.5, 243.75]	9.20	23
3	IN	> 243.75	3.36	7
4	OF19	> 22.5	10.96	30

Table 15: Groupings of DO observations in the SE Hold Pad decision tree model.

Group	Site	SE Glycol Lag	EW	Median DO	# Observations
1	IN	≤ 12775.875	≤ –0.795	9.79	21
2	IN	≤ 12775.875	(–0.795, 0.369]	3.98	6
3	IN	≤ 12775.875	> 0.369	7.88	29
4	IN	> 12775.875		3.36	2
5	OF19			9.71	58

Table 16: Groupings of COD observations in the Taxiway WK decision tree model.

Group	Site	WK Glycol	Mean COD	# of Observations
1	REF, DNST, OUT		55.4	30
2	IN	≤ 2324	262.2	9
3	IN	> 2324	920.0	1
4	OF19	≤ 2324	3321.1	9
5	OF19	> 2324	38000.0	1

Table 17: The important variables identified by the decision tree models for each deicing location (NS = north-south wind component).

Ranking	Taxiway EKS	Taxiway WK	Taxiway HY	Taxiway Z	Taxiway C	SE Hold Pad	SW Hold Pad	NE Hold Pad
1	NS	NS	NS	NS	NS	NS	NS	NS
2	Site	Glycol Usage	Glycol Usage	Site	Glycol Usage	Glycol Usage	Site	Site
3	Water Temp.	Site	Site	Water Temp.	Site	Site	Water Temp.	Water Temp.
4		Water Temp.	Water Temp.		Water Temp.	Water Temp.		