Data-Driven Optimization for Minimizing the Environmental Impact of Airport Deicing Activities

Huiyuan Fan[1*], Prashant K. Tarun[2], Victoria C.P. Chen[3], Dachuan T. Shih[4],
Jay M. Rosenberger[3], Seoung Bum Kim[5], Dan Bergman[6]

[1]Rolls-Royce Energy Systems Inc.
105 North Sandusky Street
Mount Vernon, OH 43050, USA

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

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

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

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

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

[*]Corresponding author
Tel.: +1-740-3938519
Fax: +1-740-3938336
E-mail: huiyuanfan@yahoo.com

COSMOS Technical Report 12-06

Abstract

This paper presents a data-driven deicing management framework to minimize the environmental impact of airport deicing activities at Dallas-Fort Worth International Airport. Airplane deicing involves the use of aircraft deicing and anti-icing fluids with a high concentration of ethylene/propylene/diethylene glycol. The glycol shearing off the airplanes during taxiing and dripping to the runway runs off and mixes with the airport receiving waters, which causes an increase in bacterial growth and a subsequent reduction in dissolved oxygen (DO) endangering the aquatic lives in the receiving waters. While reducing the amount of deicing and anti-icing fluid or carefully selecting runways might be more direct ways of reducing the impact of deicing activities on DO, the airport does not control those decisions. Rather, the only decision controlled by the airport is the assignment of airplanes to deicing pads. The choice of deicing pad is indirectly related to the choice of runway since each deicing pad had different runway probability distributions. The proposed deicing management framework uses stochastic dynamic programming (SDP) to assign airplanes in each hour to deicing pad locations, so as to maximize DO concentrations in the receiving waters, subject to airport constraints. The SDP formulation initially had over three hundred state variables consisting of water quality variables at six monitoring sites of the receiving water system, glycol usage variables for each of the airport's eight deicing pad locations, and number of airplanes deiced at each deicing pad location-runway combination. The state dimensionality was significantly reduced using a data mining process. The state transition equations in the SDP formulation were obtained using a two-phase statistical analysis based on artificially-generated and actual hourly data for the deicing activities. In particular, the hourly data were used to fit (i) decision tree models for the variables related to deicing activities, and (ii) multiple linear regression models for water quality variables and meteorological variables. The proposed deicing management framework was demonstrated using three cases during major deicing events. Improvements in DO compared with actual DO recorded in the data were mixed; however, the results show a promise in tackling such a complex real world problem.

Keywords: Data-driven optimization, Stochastic dynamic programming, Airport deicing, Water quality, Glycol, Dissolved oxygen

1. Introduction

Extreme winter causes ice formation on the airplanes threatening its operational safety. Therefore, deicing and anti-icing operations are performed on airplanes with utmost care during harsh winter conditions (Corsi et al. 2006, FAA Report 1996, Leist et al. 1997, Revitt and Worrall 2003, Revitt et al. 2001, Switzenbaum et al. 1999). The aircraft deicing and anti-icing fluids (ADAF) typically used in airplane deicing/anti-icing contain a high concentration of ethylene/ propylene/diethylene glycol and other proprietary additives. ADAF adhering to airplane after deicing/anti-icing can trickle down to the ground or shear off during take-off, which can run off to receiving surface waters or into the groundwater system resulting in adverse environmental impacts. Due to a high biochemical oxygen demand (BOD) and chemical oxygen demand (COD) of glycols in ADAF, the dissolved oxygen (DO) level in the receiving waters decreases (Corsi et al. 2006; U.S. Environmental Protection Agency 2000).

Every winter season Dallas/Fort Worth International Airport (D/FW Airport) experiences sporadic deicing periods requiring airplanes to be deiced/anti-iced in accordance with Federal Aviation Administration's safety regulations. Deicing practices at D/FW Airport garnered a great deal of attention in the wake of an ecological disaster in 1999 when a significant amount of glycol flowed into Trigg Lake that receives D/FW Airport runoff and subsequently killed the fish in the lake. D/FW Airport responded by upgrading its ADAF collection facilities, which included construction of eight locations equipped with deicing source isolation pads (deicing pads) where ADAF runoff is streamed into the airport's reverse osmosis wastewater treatment system. This system captures about 80% of the ADAF runoff, and the remaining 20% of the ADAF runoff occurs during taxiing and take-off that may still discharge into local receiving waters and impact both water quality and aquatic life adversely. To combat this, the airport took necessary measures elaborated in Fan et al. 2011. Furthermore, to monitor the water quality in the surrounding waterways DFW Airport joined forces with the United States Geological Survey (USGS) to set up nine monitoring sites at various locations for collecting the water quality data. A more detailed layout of the monitoring sites around the airport and the deicing pad locations at the airport can be found in Fan et al. 2011.

This research develops a data-driven optimization framework for minimizing the environmental impact of deicing activities at D/FW Airport. The framework seeks to accomplish this by maximizing expected DO concentrations in the receiving waters surrounding the D/FW Airport over a given time horizon. Alternatively, the framework could also minimize BOD or COD as objectives. However, the 1999 fish kill event was primarily the result of low DO, making it a priority for the airport. In addition, BOD and COD collection requires manual sampling and testing, as opposed to the continuous sampling (every 15-20 minutes) of DO via sensors. It should be noted that a relationship between BOD and DO exists in theory (Masters 1997, Corsi et al. 2006).

To our knowledge, there is no health-based guideline threshold value for which DO should stay below. The highest DO values observed on airport grounds occur in waterways that are not affected by airport or urban activities. The influence of airport and urban activities on DO tends downward. For simplicity, we chose to maximize the total DO since higher total DO corresponds to reduced influence of airport activities on DO. At the time of this study, the airport desired consideration of six monitoring sites in the waterways immediately surrounding the airport. In future work, the airport will focus on a few problematic waterways and penalties will be implemented to ensure sufficient DO is maintained.

To reduce the impact on DO, the amount of glycol used could be reduced or runways could be strategically selected, depending on wind conditions. However, the airport controls neither of those decisions. Rather the pilot controls the glycol usage decision, and the Federal Aviation Administration controls the runway decision. To indirectly affect the runway decision, the airport can control the assignment of airplanes to the various deicing pad locations. The different deicing pad locations have different runway probability distributions. While this setup is not ideal, the airport is the responsible party for protecting airport waterways.

A myopic framework could simply optimize each decision independently. However, the impact on DO is time-lagged, and the deicing pad locations each have limited capacity; hence, our framework utilizes stochastic dynamic programming (SDP) to enable constrained decisions that account for the future impact on DO. This SDP formulation involves over three hundred state variables, consisting of water quality variables at six monitoring sites in the airport's receiving waters, glycol usage variables for each of the airport's eight deicing pad locations, and the number of airplanes deiced at each deicing pad location-runway combination. The resulting SDP formulation apparently has the largest number of state variables in the current literature. While the optimization framework is quite generic in its applicability, our optimization prototype demonstrates results for three example cases. The method presented in this paper may help improve upon the current deicing practices by providing appropriate guidelines on *where* to deice an airplane given the meteorological conditions, deicing pad traffic, and water quality in the airport's receiving waters. The goal of this study, however, is not to dictate how deicing activities should take place at D/FW Airport, but to explore more environmental-friendly options.

2. An Overview of the Model Features

Figure 1 shows the outline of the real system and the optimization model to be built for this system. Assume that a decision is being made on an hourly basis. In time period (i.e. *stage*) 't', all airplanes to be deiced are assigned to eight deicing pad locations, DP_i, i = 1, 2, …, 8, for deicing. The six monitoring sites, MS_j, j = 1, 2, …, 6, in the airport's receiving waters help monitor the local DO levels impacted by the airport deicing activities. Actually there are nine monitoring sites set up in the airport's receiving waters, but only six sites were with the DO data and hereby are considered for our study. There are sixteen runways, RW_k, k = 1, 2, …, 16, in D/FW Airport available for deiced airplanes to take off. Taxiing and take-off of airplanes after deicing allows the ADAF to discharge into the airport's receiving waters due to drip and shear.

The *objective* of the optimization problem specifies the primary goal, which in this study is to *maximize* DO. This objective is subject to various airport *constraints* (deicing pad capacities, number of planes to be deiced). The *decision space* (or *action space*) holds the set of possible decisions or actions, and the *state space* holds the set of variables needed to describe the *state of the system*, where the system in our case is the *deicing activities system*. Uncertainty is represented by *probability distributions*. An *action* takes place whenever we need to make a *decision*. In this framework, a decision is made at the beginning of each *hour*, in which the airplanes that need to be deiced in the upcoming hour are assigned to deicing pad locations. Since the D/FW Airport operates from 5:00 AM to 11:00 PM, there are 18 stages in the SDP model indicating each hour of operation. In every hour, the state of the system must be updated. The mapping from the current state to the next state will be specified by a *state transition equation*, which is based on the statistical models. A similar data-driven approach was utilized in a previous SDP air quality research (Yang et al. 2007 and Yang et al. 2009).

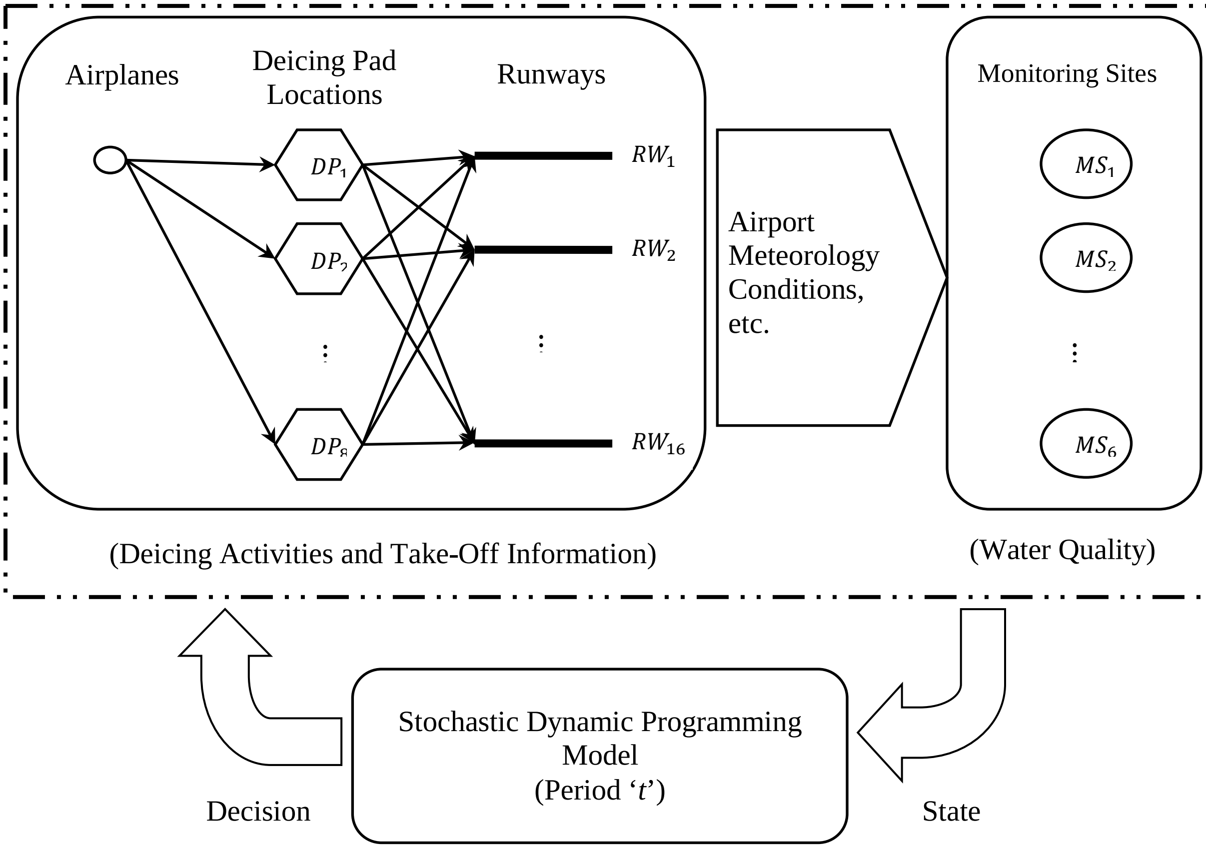

Figure 1. Schematic representation of the real system and the model to be built

3. Stochastic Dynamic Programming (SDP) Model Formulation

The objective of our optimization problem is to maximize the sum of the expected hourly-average DO concentrations over the 18 stages indicating each hour of operations. The stochastic dynamic programming (SDP) model for stage t can be formulated as below:

$$\begin{aligned} &V_t(x_t) = \max_{a_t} E\{DO_t(x_t, a_t, \varepsilon_t; u_t, w_t) + V_{t+1}(x_{t+1})\} \\ &\text{s.t.} \quad x_{t+1} = f_t(x_t, a_t, \varepsilon_t; u_t, w_t) \end{aligned} \tag{1}$$

where t = 1, 2, …, 18 correspond to deicing hours 6^{th} (5:00-6:00 a.m.), 7^{th} (6:00-7:00 a.m.), …, 23^{rd} (10:00-11:00 p.m.). In the SDP formulation above, $V_t(\cdot)$ is the future value function; x_t and a_t are state and decision vectors consisting of the state and decision variables, respectively; ε_t characterizes the stochastic nature of the system which is modeled by certain probability distributions; $DO_t(\cdot)$ is an objective function representing the DO averaged over the six monitoring sites; and $f_t(\cdot)$ is the state transition equation. Vectors u_t and w_t represent some deicing variables and external parameters that are used for modeling the objective function and the state transition equation, but are not considered as state variables.

To formulate the above SDP model for the D/FW Airport deicing system, the following additional assumptions are made:

- At the beginning of each hour, the number of airplanes to be deiced in the hour and the runway assignment for each airplane are known in advance. This is reasonable assumption given that the airplanes are present at the airport far in advance of deicing decisions, and the movement of airplanes is carefully orchestrated.
- The meteorology is not considered to be a part of the state space. It is assumed to be a known external information up to the current hour. This is also a reasonable assumption.
- The same hourly state transition equation is employed for every hour of the day. This is a simplifying assumption, and future work will model nonstationary state transition functions.
- A set of probability distributions, dependent only on meteorology, is used to estimate the amount of glycol applied to an airplane (See next Section 3.1.2). This ignores the size of the airplane as this information was not available. Future work will consider airplane size.
- As no runway information has been provided, a fixed probability distribution is used to estimate the runway used given the deicing pad location (See next Section 3.1.3). Actual runway information will be provided by the airport for future work.
- The hourly capacity of a deicing pad location is based on the assumption that the time it takes to deice an airplane is constant. This also ignores the size of the airplane. Future work will build probability distributions for the service times based on actual data.
- Only up to two time lags are used if a variable's history is considered in the SDP model. The autocorrelation and partial autocorrelation analyses presented in Fan et al. (2011) support the use of only 2 time lags of history. Future work will consider the entire history of the day in constructing nonstationary state transition functions.

A complete list of notations for the variables involved in the SDP model is given below:

Abbreviations:

DO	=	Dissolved Oxygen Concentration	[mg/l]
WT	=	Discharge	[m^3/s]

DC	=	Water Temperature	[°F]
GA	=	Glycol Amount	[kg]
GAPA	=	Glycol Amount per Airplane	[kg]
NAP	=	Number of Airplanes	
TNAP	=	Total Number of Airplanes	
AT	=	Air Temperature	[°F]
PP	=	Precipitation	[mm]
NSWS	=	North-to-South Wind Speed	[m/s]
EWWS	=	East-to-West Wind Speed	[m/s]
WD		Wind direction taking binary values (0—north-to-south; 1—south-to-north)	

Subscripts:

DP_i	=	(i = 1, …, 8) the deicing pad locations
MS_j	=	(j = 1, …, 6) the monitoring sites
RW_k	=	(k = 1, …, 16) the runways
t	=	the current stage
t-1	=	the lag one of the current stage
t-2	=	the lag two of the current stage
DP_iRW_k	=	from the deicing pad location DP_i to the runway RW_k

3.1. Actual Data and the Artificial Data Indicating Stochastic Nature of the Model

3.1.1. Actual Data

An MS-ACCESS database was created for this study that contained six groups of data actually collected from different sources. Though the data from all six groups were dealt with during the previous data mining phase with regard to its potential impact on the optimization process, the variables finally utilized in the SDP model were from the following three groups: (i) *D/FW and USGS Continuous Monitoring at 9 Sites*: DO, discharge, water temperature, precipitation, etc; (ii) *Airport Deicing Activities*: number of airplanes, ethylene and propylene glycol usage, deicing pad usage, etc.; and (iii) *Airport Meteorology*: hourly air temperature, precipitation, dew point, wind speed/direction, etc. More details on the MS-ACCESS database and the previous data mining phase can be obtained from Fan et al. (2011).

3.1.2. Modeling the Glycol Amount per Airplane as a Random Variable

The *Airport Deicing Activities* group contains daily number of airplanes. However, our SDP model is hourly based. This requires using the glycol amount per airplane as an intermediate variable to estimate the glycol usage for a deicing pad location as a potential state variable. Because of a lack of the actual hourly data, the *glycol amount per airplane* is defined as a random variable exhibiting the stochastic nature of the model. This variable was modeled by disaggregating the daily data approximately into hourly operational time slots.

The disaggregation process first used a data set provided by D/FW Airport showing the hourly number of airplanes departing from the airport during a typical operation day to obtain the fractions of airplanes departing in each hour of operation. These fractions were further used to disaggregate the daily number of airplanes deiced into hourly operational time slots. The number of airplanes for each operational day recorded in the data set was disaggregated by this process using the corresponding glycol usage data. This resulted in a sample data set for the amount of the glycol per airplane, denoted as $GAPA_t$ for the current hour *t*.

A decision tree analysis was performed on a data set obtained by matching the $GAPA_t$ data with corresponding meteorological data in order to examine the influence of meteorological variables on the amount of the glycol per airplane. The results indicated that the only meteorological variable that apparently affected the glycol amount per airplane was air temperature, AT_t. Furthermore, the decision tree split the data into 3 homogeneous groups as shown in Table 1. Groups 1, 2, and 3 have 22, 34, and 657 observations, respectively. The resulting probability distributions of glycol amount per airplane for groups 1 and 2 are shown in Tables 2 and 3, respectively. On the other hand, a Lognormal distribution of glycol amount per airplane was fit to the group 3 data with a mean of μ = 3.1363 kg and a variance of $\sigma^2 =$ 0.7621 kg^2. Given the air temperature in the hour an airplane needs to be deiced, AT_t, the appropriate group will be identified, and a realization of the glycol amount per airplane, $GAPA_t$ will be sampled from the corresponding probability distribution. This strategy is similar to the two-phase method discussed in detail in Subsection 3.5.

Table 1. Homogeneous groups for glycol amount per airplane.

Group	AT_t (in ^{o}F)
1	≤ 25.5
2	(25.5, 27.5]
3	> 27.5

Table 2. Histogram results for Group 1

Bin Interval of $GAPA_t$	Frequency

in kg	
[0, 25]	4
(25, 50]	2
(50, 100]	1
(100, 125]	2
(125, 150]	4
(150, 175]	5
(175, 200]	3
(200, 250]	1
>250	0

Table 3. Histogram results for Group 2

Bin Interval of $GAPA_t$ in kg	Frequency
[0, 20]	5
(20, 40]	7
(40, 60]	6
(60, 80]	3
(80, 100]	6
(100, 120]	4
(120, 140]	3
>140	0

3.1.3. Estimating the Runway Selection Probabilities

After deicing an airplane at a particular deicing pad location, it needs to move to a runway for takeoff. No data were available for the runway selections. At D/FW Airport, eastbound flights are assigned to runways on the east side and vice-versa for westbound flights. We had planned to incorporate destination information in our optimization. However, the airport instructed us to proceed with our optimization as a "what-if" analysis with the focus on DO. It is important to keep in mind that the results of the optimization are not intended to dictate airport decisions, rather the intention is to provide additional guidance. There are eight deicing pad locations and sixteen runways counting both directions for each of the eight runways at D/FW Airport.

Since no data were available for the runway selections, an intuitive measure of the distance between the deicing pad location and the runway was used to determine the probability that an airplane uses a particular runway after being deiced at a particular deicing pad location. As a result, we obtained a deicing pad location-runway matrix based on the premise that the smaller the distance between a deicing pad location and a runway, the higher the probability value assigned to that deicing pad location-runway combination. In other words, the probability of using a runway after being deiced at a deicing pad location is inversely proportional to the distance between the pad location and the runway. We use a scale similar in philosophy to the Analytic Hierarchy Process (AHP) (Saaty 1980) to set the probability values in the deicing pad-runway matrix for a particular wind direction. Table 4 describes the scale with the corresponding weights:

Table 4. Description of the scale with corresponding weights

Category for Distance between Deicing Pad Location and Runway	Weight
Distance Rating 1 (Lowest)	0.4
Distance Rating 2	0.3
Distance Rating 3	0.2
Distance Rating 4	0.15
Distance Rating 5	0.10
Distance Rating 6	0.075
Distance Rating 7	0.05
Distance Rating 8 (Highest)	0.025

Having assigned the weight values according to Table 4, the next step in AHP is normalization for each deicing pad location across the runways to obtain the probability that an airplane uses a particular runway after being deiced. In summary, the three steps used to generate the runway probability distribution are: 1) assign the distance rating, 2) assign the corresponding weight values according to the above scale, and 3) normalize the results in Step 2. Table 5 presents the probability values assigned to deicing pad location-runway combinations with respect to wind directions.

Table 5. Probability of using a runway after being deiced at a particular deicing pad location

	RW_1	RW_2	RW_3	RW_4	RW_5	RW_6	RW_7	RW_8
	South-to-North Wind (WD = 0)							
DP_1	0.038	0.302	0.038	0.226	0.151	0.057	0.075	0.110
DP_2	0.038	0.077	0.058	0.154	0.115	0.019	0.231	0.310
DP_3	0.019	0.154	0.038	0.308	0.231	0.058	0.077	0.120
DP_4	0.038	0.231	0.019	0.154	0.115	0.308	0.058	0.080
DP_5	0.058	0.077	0.019	0.154	0.115	0.038	0.231	0.310
DP_6	0.115	0.038	0.158	0.077	0.058	0.019	0.308	0.230
DP_7	0.038	0.077	0.058	0.154	0.115	0.019	0.231	0.310

DP_8	0.032	0.190	0.048	0.254	0.190	0.063	0.095	0.130
	RW_9	RW_{10}	RW_{11}	RW_{12}	RW_{13}	RW_{14}	RW_{15}	RW_{16}
	North-to-South Wind (*WD* = 1)							
DP_1	0.058	0.154	0.038	0.308	0.231	0.019	0.077	0.120
DP_2	0.239	0.045	0.239	0.090	0.060	0.030	0.119	0.180
DP_3	0.115	0.308	0.038	0.231	0.154	0.019	0.058	0.080
DP_4	0.058	0.038	0.019	0.308	0.231	0.077	0.115	0.150
DP_5	0.154	0.019	0.058	0.115	0.077	0.038	0.231	0.310
DP_6	0.235	0.044	0.235	0.088	0.059	0.044	0.176	0.120
DP_7	0.302	0.038	0.113	0.075	0.057	0.038	0.151	0.230
DP_8	0.195	0.195	0.049	0.195	0.146	0.049	0.073	0.100

3.2. State Variables

The state of the system at a given stage of our SDP model must include the relevant history of water quality at the six monitoring sites (hourly-averaged DO, water temperature, discharge rate), glycol usage for each deicing pad location, and the number of airplanes deiced by deicing pad location-runway combination. The number of state variables determines the dimensionality of the SDP problem, which needs to be smaller for computational reasons. For a given stage (or hour), the number of state variables is computed as below:

(6 sites)*(3 water qual. vars.) + (8 pad loc. glycol) + (8 pad loc. # airplanes)*(16 runways) = 154.

Assuming a 2-hour history (lag 1 and lag 2) and adding a state variable for the number of airplanes to be deiced in an upcoming hour, the state vector would contain 2*154 + 1 = 309 state variables. This is among the largest SDP problems in the literature, and is comparable to the air quality SDP problem of Yang et al. (2007) and Yang et al. (2009) that began with 524 state variables. In both cases, a data mining phase was used for reducing the problem dimensionality. For the air quality problem, the number of dimensions was reduced to 25 state variables. For the deicing activities optimization framework, the number of dimensions has been reduced to 45 state variables making it the largest SDP problem in the current literature. This problem is considerably larger than the current record of a 30-dimensional problem solved by Cervellera et al. (2006). For the current deicing optimization framework, the 45 state variables resulting from a previous data mining phase (Fan et al. 2011) are listed below.

- History of water quality at monitoring site, MS_6 (4 variables): $WT_{MS_6,t-1}$, $WT_{MS_6,t-2}$, $DC_{MS_6,t-1}$, $DC_{MS_6,t-2}$.
- History of DO at six monitoring sites (12 variables): $DO_{MS_j,t-1}$, $DO_{MS_j,t-2}$ for *j* = 1, …, 6.
- History of glycol usage at seven deicing pad locations (13 variables): $GA_{DP_i,t-1}$ for *i* = 1, …, 7

and $GA_{DP_i,t-2}$ for i = 1, 2, 4, 5, 6, 7.

- History of number of deiced airplanes by pad location and runway combinations (15 variables): $NAP_{DP_iRW_k,t-1}$ for (i, k) = {(1,12), (2,4), (2,7), (3,2), (4,2), (4,5), (5,7), (7,9)}, $NAP_{DP_iRW_k,t-2}$ for (i, k) = {(1,12), (2,4), (2,7), (3,2), (4,2), (4,5), (7,9)}.
- Total number of airplanes to be deiced (1 variable): $TNAP_t$

The state variables primarily consist of two lags, lag 1 and/or lag 2, for water temperature and discharge at the sixth monitoring site, DO concentrations at all six monitoring sites, glycol usages at seven of the eight deicing pad locations, and number of deiced airplanes by pad location- runway combination. For most cases, lag 1 and lag 2 for a variable show up together, with two of them, $GA_{DP_3,t-1}$ and $NAP_{DP_5RW_7,t-1}$, being exceptions with only lag 1. Of particular interest is the impact of glycol usage state variables and the runway-related state variables. The intricate relationship of the objective function and the state transition equation with decision variables is embedded in these variables, and is elaborated in the subsections next. The stochastic nature of the system can be seen in the use of the "best guessed" *ad hoc* probability distributions for sampling "data" for these variables. Consequently, these data are not guaranteed to represent reality.

3.3. Decision Variables and Constraints

There are eight decision variables in the current hour denoted as $NAP_{DP_i,t}$ for i = 1, …,8, one for each deicing pad location. Given the number of airplanes that need to be deiced in the upcoming hour, each decision variable specifies the number of airplanes assigned to a deicing pad location. The deicing pad locations have the hourly capacities as shown in Table 6. Each decision variable must be positive and with an upper limit determined by the capacity of the corresponding deicing pad location. In addition, the total number of airplanes must be equal to the sum of the decision component, i.e. $TNAP_t = \sum_{i=1}^{8} NAP_{DP_i,t}$.

Table 6. The hourly capacities of deicing pad locations

Pad location	DP_1	DP_2	DP_3	DP_4	DP_5	DP_6	DP_7	DP_8
Capacity (number of airplanes)	6	6	8	16	22	6	8	6

3.4 Additional Deicing Variables and External Parameters

Several deicing-related variables (generated using the glycol usage or runway selection probability distributions) and external parameters (mainly, for meteorology) are needed to model the objective function and the state transition equation, but they are not maintained as state variables.

These additional deicing variables are affected by the decision in the current hour and are not part of the state history. Specifically, the assignment of airplanes to deicing pad locations determines the number of airplanes being deiced at each pad location; which in turn affects the total glycol usage at each pad location and the subsequent runway selections. These 20 key variables are listed below.

- Glycol usage in the current hour at deicing pad locations (7 variables): $GA_{DP_i,t}$ for i = 1, …, 7.
- Number of deiced airplanes in the current hour by pad location and runway (13 variables): $NAP_{DP_iRW_k,t}$ for (i, k) = {(1,2),(1,12), (2,4), (2,7), (3,2), (3,4), (4,2), (4,5), (4,6), (4,12), (5,7), (5,14), (7,9)}.

The external parameters include the binary variable for wind direction, *WD*, mentioned previously. The other external parameters are meteorological variables in the current hour and their histories (lag one and lag two). These 13 meteorological variables are listed below.

- Wind direction (1 variable): *WD*
- Air temperature (3 variables): AT_t, AT_{t-1}, AT_{t-2}.
- Precipitation (3 variables): PP_t, PP_{t-1}, PP_{t-2}.
- Wind speed (6 variables): $NSWS_t$, $NSWS_{t-1}$, $NSWS_{t-2}$, $EWWS_t$, $EWWS_{t-1}$, $EWWS_{t-2}$.

3.5. State Transition Equation

The state transition equation in Equation (1) determines how each state variable changes from one hour to the next. In this model, the state variables are with one or two time lags in the currently selected hour’s water temperature and discharge at the sixth monitoring site, DO concentrations at all six monitoring sites, glycol amounts at seven deicing pad locations, and number of deiced airplanes by pad location-runway combination. When the current hour’s variables are determined, a state transition equation can be obtained as follows: the state variables with lag one are transitioned from the corresponding current hour’s state variables and the state variables with lag two are transitioned from the corresponding lag one variables. Consequently, the state transition equation for the SDP model is actually defined by the transitions for certain state variables in the current hour. Without loss of generality, these models for state variables are referred to as *state transition equations*. For these to be state transition equations, these equations should be able to predict the current hour’s variables based on past histories and the current hour's decision. In particular, the state transition equations that include decision variables are quite significant for the optimization process.

Unfortunately, for the glycol amount related state variables and the runway-related state variables described in Subsection 3.2 due to a lack of the actual data, their current hour's state variables were predicted using probability distributions and hence no past histories could be considered. Nevertheless, the decision in the current hour is utilized in estimating those variables. Once the deicing variables discussed in Subsection 3.4 are determined, the required transitions of these state variables can be estimated. For the variable total number of the airplanes deiced, $TNAP_t$, transition is not required. The remaining state variables are water temperature and discharge at the sixth monitoring site, and the DO concentrations at all six monitoring sites. The current hour's values for these variables are predicted by statistical models built from the actual data and artificially generated data discussed previously. Furthermore, each of these state transitions is a *function* of *water quality variables* from hours *t*–1 and earlier, and *decision variables* from hours *t* and earlier. These equations, therefore, form the core of the state transitions for this optimization framework.

For developing the state transition equations, the analysis was conducted in two phases combining decision tree and linear regression concepts, as in treed regression (Alexander and Grimshaw 1996, Kim et al. 2007):

- The *first phase* uses decision tree models to incorporate the *deicing activity variables*: glycol amounts at each deicing pad location and number of planes by deicing pad location-runway combination. This phase helps the modeling process by decomposing a space into certain subspaces based on the terminal nodes in the decision trees, which facilitate modeling in the second phase.
- The *second phase* refines the approximation from the first phase using stepwise linear regression on the *water quality* and *meteorological variables* for the terminal node data from the decision trees of the first phase. Only the terminal nodes with sufficient data were used in this phase. Those terminal nodes with too few (less than 10) observations used the average value as the predicted value.

The primary reason for this two-phase approach was to provide better opportunity for the deicing activity variables to appear in the models. If, by some chance, no deicing variables appear in the models, then the optimization would have no means to control the system. In both phases, only the important variables identified from a previous data mining phase are employed to build the equations, and only the variables identified as important by the tree modeling and stepwise regression procedures are maintained in the equations. These state transition equations include state variables, deicing variables and external variables specified in Subsections 3.2, 3.3, and 3.4. CART software (www.salfordsystems.com) was used to construct the *first phase* decision trees, and SAS software (www.sas.com) was used to build the *second phase* multiple linear regression models.

In this section, we discuss the DO model for the first monitoring site in detail for demonstration purposes. Only a brief summary of the other models is given. The *first phase* decision tree model for $DO_{MS_1,t}$ at the monitoring site MS_1 resulted in 8 terminal nodes representing 8 homogeneous groups, with split points specified in Table 7. In other words, the input space of $DO_{MS_1,t}$ was decomposed into 8 subspaces. The *second phase* linear regression models and corresponding R^2 values (where values close to 1.0 indicate excellent fit) for the homogenous groups are shown in Table 8. Note that groups 2 and 7 were not used in the construction of the *second phase* linear regression models because there were too few observations to come up with a reliable linear regression model.

Table 7. Homogeneous groups in the first phase decision tree for DO at the first monitoring site

Group	1	2	3	4	5	6	7	8
$GA_{DP_5,t-2}$	≤ 1.71	≤ 1.71	≤ 1.71	≤ 1.71	≤ 1.71	≤ 1.71	≤ 1.71	>1.71
$GA_{DP_1,t}$	≤ 2.36	≤ 2.36	≤ 2.36	≤ 2.36	≤ 2.36	≤ 2.36	≤ 2.36	
$NAP_{DP_4RW_5,t-2}$						≤ 0.5	>0.5	
$GA_{DP_1,t-1}$	≤ 2.36	≤ 2.36	≤ 2.36	≤ 2.36	> 2.36			
$GA_{DP_1,t-2}$	≤ 35.05	≤ 35.05	≤ 35.05	> 35.05				
$GA_{DP_6,t-2}$	≤ 14.64	≤ 14.64	> 14.64					
$GA_{DP_5,t}$	≤ 5.0	> 5.0						

Table 8. Second phase results for estimated state transition equations for DO at the first monitoring site.

Group	R^2	Estimated state transition equations for $DO_{MS_1,t}$ at MS_1
1	0.9855	$DO_{MS_1,t} = 0.302 - 0.00094348{*}EWWS_{t-2} - 0.00197{*}AT_t + 0.00127{*}AT_{t-2} + 1.21739{*}DO_{MS_1,t-1} - 0.23749{*}DO_{MS_1,t-2} - 0.00502{*}WT_{MS_6,t-2}$
2		$DO_{MS_1,t} = 12.653$
3	0.9997	$DO_{MS_1,t} = -0.41054 + -0.01646{*}EWWS_{t-1} - 0.00465{*}NSWS_{t-2} + 1.03229{*}DO_{MS_1,t-1} - 0.00209{*}DC_{MS_6,t-1} + 0.4778{*}WT_{MS_6,t-1} - 0.46060{*}WT_{MS_6,t-2}$
4	0.9734	$DO_{MS_1,t} = 1.00432 + 0.95152{*}DO_{MS_1,t-1} - 0.04841{*}WT_{MS_6,t-2}$
5	0.9856	$DO_{MS_1,t} = 0.15076 - 0.00367{*}NSWS_{t-2} + 1.26217{*}DO_{MS_1,t-1} - 0.27566{*}DO_{MS_1,t-2}$
6	0.9870	$DO_{MS_1,t} = 0.45669 - 0.00515{*}AT_{t-2} + 1.30756{*}DO_{MS_1,t-1}$

		$- 0.32843*DO_{MS_1,t-2}$
7		$DO_{MS_1,t} = 15.007$
8	0.9732	$DO_{MS_1,t} = 0.90857 + 1.41939*DO_{MS_1,t-1} - 0.47743*DO_{MS_1,t-2}$ $- 0.02447*WT_{MS_6,t-1}$

As can be seen in Tables 7 and 8 for the first monitoring site, the deicing activity-related variables are used to identify the group and the water quality and meteorological variables are used to predict the DO concentration in the group. It is important to note that such a model overall defines a target state variable as a function of all the variables involved in the two phase process. As a result, the model for $DO_{MS_1,t}$ can be a function of fifteen variables, including seven deicing activity-related variables, and eight water quality-related variables, as shown in Tables 7 and 8 and further listed in Table 9.

Similar models were built for DO at other five monitoring sites, and for water temperature and discharge at the sixth monitoring site. These models are also summarized in Table 9. It can be seen from the table that: (a) each model has certain state variables, decision variables, deicing variables and external parameters; (b) a current hour's variable is related with its history; and c) decision and stochastic nature are implicitly represented with the deicing activity-related variables.

Table 9. Summary of the state transition models for DO at all six monitoring sites and for water temperature and discharge at the sixth monitoring site.

State variable	**Number of homogeneous groups in the decision tree**	**Variables in the first phase**		**Variables in the second phase**	
		Glycol amounts	**Combinations of deicing pad locations vs. runways**	**Water quality variables**	**Meteorological variables**
$DO_{MS_1,t}$	8	$GA_{DP_1,t}$ $GA_{DP_1,t-1}$ $GA_{DP_1,t-2}$ $GA_{DP_5,t}$ $GA_{DP_5,t-2}$ $GA_{DP_6,t-2}$	$NAP_{DP_4RW_5,t-2}$	$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$ $DO_{MS_1,t-1}$ $DO_{MS_1,t-2}$	AT_t AT_{t-2} $NSWS_{t-2}$ $EWWS_{t-2}$
$DO_{MS_2,t}$	6	$GA_{DP_1,t}$ $GA_{DP_1,t-1}$ $GA_{DP_1,t-2}$ $GA_{DP_2,t}$		$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$ $DO_{MS_2,t-1}$ $DO_{MS_2,t-2}$	AT_t AT_{t-2}

		$GA_{DP_3,t}$			
$DO_{MS_3,t}$	7	$GA_{DP_1,t}$ $GA_{DP_1,t-2}$ $GA_{DP_4,t-1}$ $GA_{DP_6,t}$ $GA_{DP_6,t-2}$ $GA_{DP_7,t-1}$		$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$ $DC_{MS_6,t-2}$ $DO_{MS_3,t-1}$ $DO_{MS_3,t-2}$	AT_t AT_{t-2} $NSWS_t$ $NSWS_{t-1}$ $EWWS_{t-1}$
$DO_{MS_4,t}$	11	$GA_{DP_1,t-1}$ $GA_{DP_3,t-1}$ $GA_{DP_6,t-1}$ $GA_{DP_6,t-2}$ $GA_{DP_7,t-1}$	$NAP_{DP_1RW_2,t}$ $NAP_{DP_3RW_2,t-1}$ $NAP_{DP_3RW_2,t-2}$ $NAP_{DP_3RW_4,t}$ $NAP_{DP_4RW_{12},t}$	$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$ $DC_{MS_6,t-2}$ $DO_{MS_4,t-1}$ $DO_{MS_4,t-2}$	AT_t AT_{t-1} AT_{t-2} PP_t $NSWS_{t-1}$ $NSWS_{t-2}$ $EWWS_{t-2}$
$DO_{MS_5,t}$	14	$GA_{DP_1,t}$ $GA_{DP_1,t-1}$ $GA_{DP_1,t-2}$ $GA_{DP_4,t-2}$ $GA_{DP_6,t}$ $GA_{DP_6,t-2}$	$NAP_{DP_1RW_{12},t}$ $NAP_{DP_1RW_{12},t-1}$ $NAP_{DP_1RW_{12},t-2}$ $NAP_{DP_4RW_2,t-2}$ $NAP_{DP_4RW_5,t}$ $NAP_{DP_5RW_7,t-1}$ $NAP_{DP_5RW_{14},t}$	$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$ $DC_{MS_6,t-1}$ $DO_{MS_5,t-1}$ $DO_{MS_5,t-2}$	AT_t AT_{t-2} $NSWS_t$ $NSWS_{t-2}$
$DO_{MS_6,t}$	16	$GA_{DP_1,t}$ $GA_{DP_1,t-1}$ $GA_{DP_1,t-2}$ $GA_{DP_4,t}$ $GA_{DP_4,t-1}$ $GA_{DP_4,t-2}$ $GA_{DP_5,t}$ $GA_{DP_5,t-2}$ $GA_{DP_6,t}$ $GA_{DP_6,t-2}$ $GA_{DP_7,t}$ $GA_{DP_7,t-1}$ $GA_{DP_7,t-2}$	$NAP_{DP_2RW_7,t-2}$	$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$ $DO_{MS_6,t-1}$ $DO_{MS_6,t-2}$	AT_t AT_{t-2} PP_t $NSWS_{t-2}$
$WT_{MS_6,t}$	29	$GA_{DP_1,t}$ $GA_{DP_1,t-1}$ $GA_{DP_1,t-2}$	$NAP_{DP_1RW_{12},t}$ $NAP_{DP_1RW_{12},t-1}$ $NAP_{DP_1RW_{12},t-2}$	$WT_{MS_6,t-1}$ $WT_{MS_6,t-2}$	AT_t AT_{t-1} AT_{t-2} $NSWS_t$

		$GA_{DP_2,t}$ $GA_{DP_2,t-1}$ $GA_{DP_2,t-2}$ $GA_{DP_4,t}$ $GA_{DP_4,t-1}$ $GA_{DP_4,t-2}$ $GA_{DP_5,t}$ $GA_{DP_6,t}$ $GA_{DP_6,t-1}$ $GA_{DP_6,t-2}$ $GA_{DP_7,t}$ $GA_{DP_7,t-1}$ $GA_{DP_7,t-2}$	$NAP_{DP_7RW_9,t-1}$ $NAP_{DP_7RW_9,t-2}$		$NSWS_{t-2}$ $EWWS_t$ $EWWS_{t-1}$ $EWWS_{t-2}$
$DC_{MS_6,t}$	5	$GA_{DP_1,t-2}$ $GA_{DP_7,t-2}$		$DC_{MS_6,t-1}$ $DC_{MS_6,t-2}$	AT_t PP_t $NSWS_t$ $NSWS_{t-1}$ $EWWS_t$ $EWWS_{t-2}$

3.6. Objective Function

The state transition models for DO at six monitoring sites discussed in the previous subsection help to determine the objective function of the SDP formulation shown in Equation (1) as below:

$$DO_t = \frac{1}{6}\sum_{j=1}^{6} DO_{MS_j,t} \tag{2}$$

3.7. An Interesting Remark

It was gathered from the SDP modeling process that the eighth deicing pad, DP_8, did not affect the modeling process as its deicing activity-related variables did not appear in the list of state variables as well as in any model for objective function or for state transition. Whether it reflects the reality of D/FW Airport's deicing system is not clear at this point and can be an interesting project for future.

4. Deicing Activities Optimization Framework: Solution & Implementation Results

Three days with relatively heavier demand for deicing at the D/FW Airport (Case 1: 01/12/2003; Case 2: 02/24/2003; Case 3: 02/25/2003) during the 2003 deicing season were selected as instances to demonstrate the performance of the proposed data-driven SDP framework. The idea was to build a tool that could help optimize deicing activities at the D/FW Airport in terms of its impact on its surrounding receiving waters.

4.1. Approximate SDP Solution

The SDP formulation for the deicing activities optimization framework is high-dimensional with 45 state variables (shown in Subsection 3.2). Such complex SDP problems can only be solved approximately. We utilize the approximate SDP solution approach described in Cervellera et al. (2006). The major steps in the solution process for a given hour t are shown in Figure 2. Given a point x_t that specifies values of the state variables within the state space, the optimizer solves for a point V_t on the future value function. Sufficient (x_t, V_t) pairs are obtained, and they form the training data set for constructing a neural network approximation $\hat{V}_t(\cdot)$ of the future value function. This process of obtaining (x_t, V_t) pairs is controlled by design of experiments (DOE). For each of the three prototypical cases mentioned previously, we randomly sampled 200 x_t points in the state space, and used a neural network with one hidden layer containing 20 nodes. Once the future value function approximations $\hat{V}_t(\cdot)$ are obtained, the optimal decisions can be generated for any state of the system. We wrote the SDP code in Matlab software (www.mathworks.com) using Optimization and Neural Network Toolboxes to solve for 18 future value function approximations (one for each hour), then generated optimal decisions for specific cases using our MATLAB simulation model.

4.2. Special Considerations

Some special considerations for the three prototypical cases include:

- For each case, the SDP problem was solved using actual meteorological data and the actual number of airplanes deiced.
- As no information was available for the daily take-off directions for the three cases, the take-off direction for a day was based on the wind direction (WD) in the first operational hour of the airport on that day.

- Because the eighth deicing pad does not affect the DO for any monitoring sites, we used this result to reduce the decision space dimension by setting up a rule that made it the most preferred location choice to deice airplanes.
- Basically, the probability distributions presented in the Subsection 3.1.3 for the runway selections contribute to the uncertainty of the SDP model. In this application, it was only used deterministically to assign the runways for airplanes to take off after being deiced at given pad location. In other words, it was not included in the expectation value estimation in Equation (1) over the sampled data points at certain times.

4.3. Optimization Results and Discussions

The solution runtimes for the SDP formulation using the MATLAB Software on a Dual 3.06-GHz Intel Xeon Workstation were 48.25 hours for Case 1, 42.5 hours for Case 2, and 38 hours for Case 3. Each case was simulated using the optimal decisions, and the results are shown in Tables 10 through 15. The results in Tables 10, 12, and 14 are simulated realizations of the SDP policy. These are shown for illustration purposes since this reflects the type of result the airport would see on a specific day. The simulation employs the SDP policy and would dynamically alter the decisions depending on the realized states. The simulation runtimes were about 5 minutes for each case. The results should be viewed with an understanding that it is based on some artificially generated data and several simplifying assumptions. In addition, the MATLAB optimizer could not converge on several instances due to inherent non-convexity issues with the optimization problem of this nature. To accommodate the trees in the MATLAB optimizer, we represented the decision trees by a sequence of “if-then” rules. However, the MATLAB optimizer is still a generic routine that is not specialized to guarantee optimality in the presence of treed regression models. The issues related to trees in an optimization framework will be further examined in the future.

From the results for Case 1, it is seen that an improvement in the average DO with respect to the actual average DO based on the D/FW & USGS data occurs at the first and fourth monitoring sites, whereas the average DO at the third monitoring site is quite close to the actual DO. It is important to mention here that the D/FW Airport installed some aerators in the Trigg Lake. Geographically, the second monitoring site could have been impacted by these aerators and other urban activities that were not explicitly modeled in this optimization framework.

The results for Case 2 indicate an improvement in the average DO over the actual average DO at the fourth monitoring site, whereas the average DO is quite close to the actual DO at the first and fifth monitoring sites.

The results for Case 3 indicate an improvement in the average DO over the actual average DO at the third and fifth monitoring sites, whereas the average DO is quite close to the actual DO at the first, fourth and sixth monitoring sites.

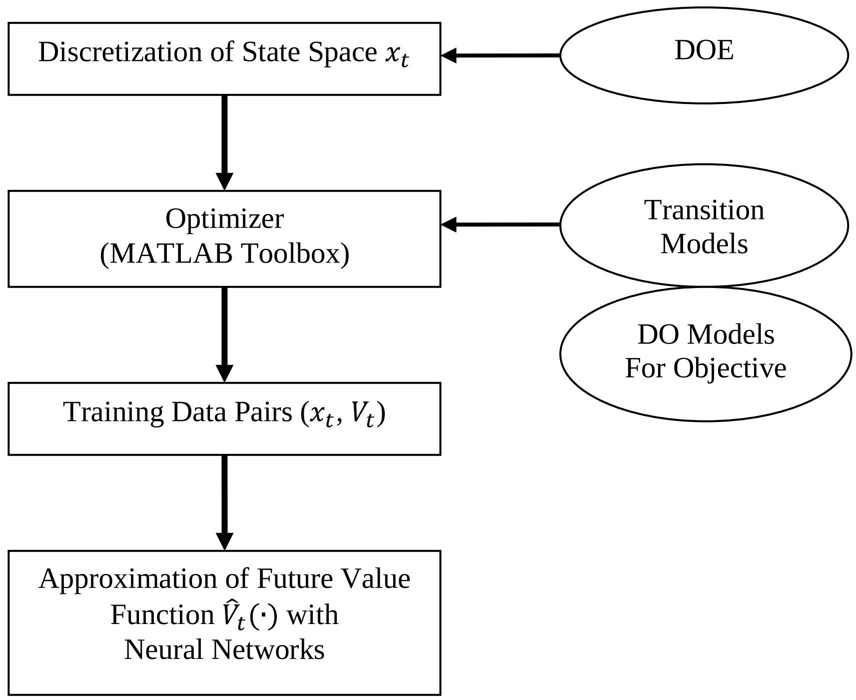

Figure 2. Flowchart of the approximate SDP approach in hour *t*.

Table 10. Case 1: The optimal deicing pad location assignments for each deicing hour.

Stage *t*	Hour	$TNAP_t$	DP_1	DP_2	DP_3	DP_4	DP_5	DP_6	DP_7	DP_8
1	6	9	0	0	2	0	0	1	0	6
2	7	43	6	0	3	12	10	6	0	6
3	8	37	0	6	8	16	1	0	0	6
4	9	54	0	4	8	0	22	6	8	6
5	10	64	0	6	6	16	22	6	2	6
6	11	72	2	6	8	16	22	4	8	6
7	12	62	6	4	8	16	16	6	0	6
8	13	75	6	6	5	16	22	6	8	6
9	14	65	6	3	8	13	22	2	5	6
10	15	58	0	6	7	16	22	1	0	6
11	16	66	6	6	8	4	22	6	8	6
12	17	60	0	6	8	16	22	2	0	6
13	18	65	0	6	8	9	22	6	8	6

14	19	70	4	0	8	16	22	6	8	6
15	20	60	6	6	2	16	22	2	0	6
16	21	47	0	6	5	0	22	0	8	6
17	22	45	6	5	0	0	20	0	8	6
18	23	0	0	0	0	0	0	0	0	0

Table 11. Case 1: The realized DO values at each monitoring site under the optimal assignments.

Stage *t*	Hour	MS_1	MS_2	MS_3	MS_4	MS_5	MS_6
1	6	11.7794	8.8572	9.7143	13.8534	8.2895	12.6529
2	7	11.7794	8.8572	9.6956	14.0625	8.2895	12.6529
3	8	12.653	7.8246	9.7252	14.0625	8.2511	12.6101
4	9	11.7794	8.8572	9.6956	13.8802	8.2895	12.6553
5	10	11.7794	8.8572	9.6956	14.0625	8.2895	12.6553
6	11	11.7794	8.8572	9.6956	14.0625	8.2895	12.6578
7	12	11.7794	8.8572	9.6956	14.0625	8.2895	12.6578
8	13	11.7794	8.8572	9.6956	14.0625	8.2895	12.6602
9	14	11.7794	8.8572	9.6956	14.0625	8.2895	12.6602
10	15	11.7794	8.8572	9.6956	14.0625	8.2895	12.6602
11	16	11.7794	8.8572	9.6956	14.0625	8.2895	12.6651
12	17	12.653	7.8246	9.6956	14.0625	12.148	12.6101
13	18	11.7794	8.8572	9.6956	14.0625	12.148	12.6724
14	19	11.7794	8.8572	9.6956	14.0625	8.2895	12.6773
15	20	11.7794	8.8572	9.6956	14.0625	8.2895	12.6773
16	21	11.7794	8.8572	9.6956	13.915	8.2895	12.6773
17	22	11.7794	8.8572	9.6956	13.922	8.2895	12.6773
18	23	11.7794	8.8572	9.7143	13.9172	8.2895	12.6773
Average		**11.8765**	**8.7425**	**9.6993**	**14.0167**	**8.7161**	**12.6587**
Actual (5AM-11PM)		11.77		9.82	12.96	9.91	16.5

Table 12. Case 2: The optimal deicing pad location assignments for each deicing hour.

Stage *t*	Hour	$TNAP_t$	DP_1	DP_2	DP_3	DP_4	DP_5	DP_6	DP_7	DP_8
1	6	2	0	0	0	0	0	0	0	2
2	7	22	6	1	7	1	0	1	0	6
3	8	26	0	0	0	0	14	5	1	6
4	9	34	6	0	0	4	7	3	8	6
5	10	44	4	0	0	10	22	0	2	6
6	11	32	6	0	0	0	20	0	0	6
7	12	39	6	6	1	8	0	6	6	6
8	13	40	6	0	0	14	0	6	8	6
9	14	38	6	0	8	3	1	6	8	6
10	15	37	0	2	8	13	0	0	8	6

11	16	36	0	0	8	0	18	4	0	6
12	17	29	0	0	2	9	4	0	8	6
13	18	57	6	6	8	0	22	6	3	6
14	19	43	6	6	8	15	1	0	1	6
15	20	36	0	6	0	16	2	6	0	6
16	21	30	0	1	2	15	0	4	2	6
17	22	22	0	6	5	5	0	0	0	6
18	23	1	0	0	0	0	0	0	0	1

Table 13. Case 2: The realized DO values at each monitoring site under the optimal assignments.

Stage *t*	Hour	MS_1	MS_2	MS_3	MS_4	MS_5	MS_6
1	6	10.9847	18.9428	10.7286	6.678	11.6265	10.6762
2	7	11.023	17.6508	10.7289	6.6681	11.6577	10.7336
3	8	11.023	17.6508	10.7289	6.6838	11.6577	10.7336
4	9	11.023	17.6508	10.7289	7.9565	11.6577	10.736
5	10	11.023	17.6508	10.7289	7.9565	11.6577	10.7385
6	11	11.023	17.6508	10.7289	6.7072	11.6577	10.7409
7	12	11.023	17.6508	10.7289	7.9565	11.6577	10.7433
8	13	11.023	17.6508	10.7289	7.9565	11.6577	10.7458
9	14	11.023	17.6508	10.7289	7.9565	11.6577	10.7458
10	15	11.023	17.6508	10.7289	7.9565	11.6577	10.7482
11	16	12.653	19.3339	10.7289	6.6995	12.148	10.7087
12	17	11.023	17.6508	10.7289	7.9565	11.6577	10.7458
13	18	11.023	17.6508	10.7289	6.6884	11.6577	10.7458
14	19	11.023	17.6508	10.7289	7.9565	11.6577	10.7433
15	20	12.653	19.3339	10.7289	7.9565	12.148	10.7087
16	21	10.9968	19.3339	10.7289	7.9565	12.148	10.7087
17	22	11.023	17.6508	10.7289	7.9565	11.6577	10.7433
18	23	10.9956	18.9595	10.761	6.6961	11.651	10.6762
Average		**11.1990**	**18.0758**	**10.7307**	**7.4635**	**11.7373**	**10.7290**
Actual (5AM-11PM)		11.67		12.12	6.65	12.15	11.36

Table 14. Case 3: The optimal deicing pad location assignments for each deicing hour.

Stage *t*	Hour	$TNAP_t$	DP_1	DP_2	DP_3	DP_4	DP_5	DP_6	DP_7	DP_8
1	6	3	0	0	0	0	0	0	0	3
2	7	15	0	3	4	0	0	0	2	6
3	8	22	0	0	8	0	4	3	1	6
4	9	31	1	1	1	1	20	0	1	6
5	10	27	0	0	0	0	7	6	8	6
6	11	27	3	0	5	3	2	5	3	6
7	12	34	3	1	0	12	0	6	6	6
8	13	41	0	1	1	15	4	6	8	6

9	14	38	3	6	8	0	5	3	7	6
10	15	26	0	0	0	2	17	1	0	6
11	16	27	0	4	1	0	16	0	0	6
12	17	30	0	6	8	0	0	2	8	6
13	18	37	6	6	8	3	0	0	8	6
14	19	37	5	6	6	0	0	6	8	6
15	20	35	1	1	1	5	21	0	0	6
16	21	23	0	0	8	0	0	6	3	6
17	22	23	0	6	0	7	4	0	0	6
18	23	0	0	0	0	0	0	0	0	0

Table 15. Case 3: The realized DO values at each monitoring site under the optimal assignments.

Stage *t*	Hour	MS_1	MS_2	MS_3	MS_4	MS_5	MS_6
1	6	13.4011	27.4926	14.0247	4.121	13.0618	13.0659
2	7	13.4011	27.4926	13.9902	4.1322	13.0618	13.0684
3	8	13.4011	27.4926	13.9902	4.1304	12.148	13.0684
4	9	13.4011	27.4926	13.9902	4.1091	13.0618	13.0684
5	10	13.4011	27.4926	13.9902	4.1152	13.0618	13.0684
6	11	13.4011	27.4926	13.9902	6.746	13.0618	13.0684
7	12	13.4011	27.4926	13.9902	6.746	13.0618	13.0684
8	13	13.4011	27.4926	13.9902	6.746	13.0618	13.0684
9	14	13.4011	27.4926	13.9902	4.124	13.0618	13.0684
10	15	13.4011	27.4926	13.9902	6.746	13.0618	13.0684
11	16	13.4011	27.4926	13.9902	4.0979	13.0618	13.0635
12	17	12.653	20.3551	13.9902	4.0762	12.148	13.0142
13	18	13.4011	27.4926	14.0247	6.746	13.0618	13.061
14	19	13.4011	27.4926	13.9902	4.0916	13.0618	13.0586
15	20	13.4011	27.4926	13.9902	6.746	13.0618	13.0562
16	21	13.4011	27.4926	13.9902	4.0717	13.0618	13.0537
17	22	12.653	20.3551	13.9902	6.746	12.148	13.0026
18	23	12.653	26.2488	13.9653	4.0695	13.0122	13.0305
Average		**13.2764**	**26.6304**	**13.9927**	**5.1312**	**12.9067**	**13.0568**
Actual (5AM-11PM)		13.96		12.67	5.3	12.89	13.3

5. Conclusions and Future Work

A data-driven deicing activities management framework to minimize the environmental impact of airport deicing activities at Dallas-Fort Worth International Airport has been developed. The framework utilizes stochastic dynamic programming (SDP) approach to maximize the dissolved oxygen (DO) at the six monitoring sites in the airport's receiving water system over 18 operating hours in a day, subject to the airport constraints. The reduced state space as a result of a data mining process includes 45 state variables consisting of water quality variables at six monitoring sites, glycol usage for each deicing pad location, and the number of airplanes deiced by deicing pad location-runway combination. The actual and artificially generated hourly data were used to conduct a two-phase statistical analysis: (i) to fit decision tree models for the variables related to deicing activities, and (ii) to fit multiple linear regression models for water quality variables and meteorological variables. Statistical models were constructed for DO at all six monitoring sites in the airport's receiving waters, and for water temperature and discharge rate at the sixth monitoring site. They served as state transition equations in the deicing activities optimization framework. Given the number of airplanes that needs to be deiced in the upcoming hour, decision variables specify the assignment of these airplanes to deicing pad locations. The proposed deicing activities optimization framework demonstrates the potential to execute the optimization process for a real system as complex as the deicing activities system at D/FW airport. Such an optimization/simulation tool can be used to study actual D/FW operational scenarios under various conditions and come up with a set of recommendations to improve the ecological impact of the current deicing practices at the airport.

Recently, the airport implemented an active data collection of deicing activities which will be important for our framework. For future work, first, the optimization will focus on only the problematic monitoring sites, and it will implement a penalty function approach that seeks to achieve DO above a desired threshold, instead of simply maximizing DO. Second, nonstationary state transitions will be modeled using the same treed regression approach, and data mining and variable selection procedures will be critical for dimensionality reduction. Third, a specialized mixed-integer optimization code will be developed to more appropriately handle the treed regression models within the optimization. Finally, a new adaptive value function approximation method (Fan et al. 2013) will be employed to solve for the SDP policy.

Acknowledgements

This research was supported by the Dallas-Fort Worth International Airport and the National Science Foundation (ECCS-0801802).

References

Alexander, W. P. and S. D. Grimshaw (1996). "Treed Regression." *Journal of Computational and Graphical Statistics*, **5**(2), 156–175.

Cervellera, C., V. C. P. Chen, and A. Wen (2006). "Optimization of a Large-Scale Water Reservoir Network by Stochastic Dynamic Programming with Efficient State Space Discretization." *European Journal of Operational Research*, **171**, 1139–1151.

Corsi, S.R., G. R. Harwell, S. W. Geis, and D. Bergman (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** (11), 2890-2900.

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)

Fan, H., P. K. Tarun, D. T. Shih, S. B. Kim, V. C. P. Chen, J. M. Rosenberger, and D. Bergman (2011). "Data Mining Modeling on the Environmental Impact of Airport Deicing Activities." *Expert Systems with Applications*, **38**, 14899–14906.

Fan, H, P. K. Tarun, and V. C. P. Chen (2013). "Adaptive Value Function Approximation for Continuous-State Stochastic Dynamic Programming." *Computers & Operations Research*, **40**, 1076-1084.

Leist, D., K. Flowers, G. Bartley, B. Beebe, T. Giles, and E. Ratcliff (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).

Masters, G. M. (1997). *Introduction to Environmental Engineering and Science* (2nd edition). Upper Saddler River, NJ: Prentice Hall.

Kim, H., W.-Y. Loh, Y.-S. Shih, and P. Chaudhuri (2007). "Visualizable and interpretable regression models with good prediction power." *IIE Transactions*, **39**, 565–579.

Revitt, D.M., and P. Worrall (2003). "Low Temperature Biodegradation of Airport De-icing Fluids." *Water Science and Technology*, **48** (9), 103-111.

Revitt, D.M., P. Worrall, and D. Brewer (2001). "The Integration of Constructed Wetlands into a Treatment System for Airport Runoff." *Water Science and Technology*, **44** (10,11), 469-476.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill.

Switzenbaum, M.S., S. Veltman, T. Schoenberg, G. M. Durand, D. Mericas, and B. Wagoner (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).

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

Yang, Z., V. C. P. Chen, M. E. Chang, T. E. Murphy, and J. C. C. Tsai (2007). "Mining and Modeling for a Metropolitan Atlanta Ozone Pollution Decision-Making Framework." *IIE Transactions*, Special Issue on Data Mining, **39**, 607–615.

Yang, Z., V. C. P. Chen, M. E. Chang, M. L. Sattler, and A. Wen (2009). "A Decision-Making Framework for Ozone Pollution Control." *Operations Research*, **57**(2), pp. 484-498.