MARYLAND DEPARTMENT OF NATURAL RESOURCES

DNR 12-10212005-69
PPRP-133

PPRP

Data Mining to Determine Local Effects of Mercury Emissions

MARYLAND POWER PLANT RESEARCH PROGRAM

November 2005

Robert L. Ehrlich, Jr., Governor ♦ Michael S. Steele, Lt. Governor

The Maryland Department of Natural Resources (DNR) seeks to preserve, protect and enhance the living resources of the State. Working in partnership with the citizens of Maryland, this worthwhile goal will become a reality. This publication provides information that will increase your understanding of how DNR strives to reach that goal through its many diverse programs.

C. Ronald Franks, Secretary
Maryland Department of Natural Resources

The facilities and services of the Maryland Department of Natural Resources are available to all without regard to race, color, religion, sex, sexual orientation, age, national origin or physical or mental disability.

This document is available in alternative format upon request from a qualified individual with a disability.

Maryland Department of Natural Resources
Tawes State Office Building
580 Taylor Avenue
Annapolis, Maryland 21401-2397
Toll Free in Maryland: 1-877-620-8DNR x8660
Outside Maryland: 1-410-260-8660
TTY users call via the Maryland Relay
www.dnr.maryland.gov

Printed on Recycled Paper

Data Mining to Determine Local Effects of Mercury Emissions

November 2005

Prepared under the direction of:

John Sherwell, Ph.D.
Maryland Power Plant Research Program

Prepared by:

Kenneth Walsh, Ph.D.
Matthew Milligan
Science Applications International Corporation
615 Oberlin Road, Suite 100
Raleigh, NC 27605

Prepared for:

Maryland Department of Natural Resources
Power Plant Research Program
580 Taylor Avenue
Tawes State Office Building, B-3
Annapolis, Maryland 21401

ACKNOWLEDGEMENTS

The paper was prepared by Science Applications International Corporation (SAIC) as subcontractor to Environmental Resources Management, Inc. (ERM) under the Maryland Department of Natural Resources Power Plant Research Program (PPRP) Atmospheric Sciences Integrator contract (#K00B0200086).

FOREWORD

The Maryland Power Plant Research Program funded this work to construct a piecewise linear regression model to describe the effect that local mercury emissions would have on ambient mercury deposition rates and concentrations. To accomplish this goal, two candidate sites were chosen from the Mercury Deposition Network that had operated for significant periods and that were located near single large coal-fired power plants, the major local sources of mercury emissions. The mercury deposition rates and concentrations, heat input for the power plant (a surrogate for emissions), and meteorology were compared to understand the various influences.

By understanding the causes for increases in mercury deposition rates, Maryland will be better able to understand the effects of local emissions sources on ambient mercury levels. This improved understanding may result in more effective strategies to protect public health.

This study also shows how a data mining technique can assess the relative roles of emissions and meteorology to describe ambient air quality over several years. Such techniques can quickly provide policy makers with behavioral information and measures of accuracy based on historical data sets. The techniques do not require the computer-intensive resources of classical modeling techniques and can be performed in seconds to establish the major parameters affecting air quality.

ABSTRACT

Tools for knowledge discovery through databases (KDD) can provide air quality planners with descriptions of long-term behavior at monitor sites. In this project, the Maryland Power Plant Research Program used the Cubist® program to construct a piecewise linear regression model to describe mercury deposition rates and concentrations at two sites in terms of hourly power plant heat input (a surrogate for emissions) and meteorological variables. The ambient mercury measurements represented more than five years of data at the Marcell, Minnesota and Chassahowitzka, Florida sites of the Mercury Deposition Network. The daily heat input data represented two power plants (Boswell Energy Center and Florida Power's Crystal River Energy Complex) that are predominantly coal-fired.

The advantages of local versus regional controls for mercury emissions have recently been debated. This work aimed to identify the meteorological conditions that lead to an effect of local emissions on local ambient mercury conditions. The long time series of ambient mercury deposition data at the Marcell and Chassahowitzka sites were compared with the simultaneous heat input quantities at the two nearby power plants.

Meteorological conditions were included in the data sets, and the Cubist software generated piecewise linear regression fits that describe the deposition data based on heat input and meteorological parameters. The software created models of several linear expressions that apply under particular meteorological conditions and sometimes overlap. The investigation showed that the effect of power plants on nearby mercury deposition sites only manifested itself under particular meteorological conditions. The data analyses and findings are discussed in detail.

The investigation pointed to a rise in mercury deposition at the Marcell site associated with rate increases in heat input at the Boswell Energy Center, likely occurring during boiler startup or shutdown. The spikes above the average deposition represent only eight percent of the weekly data at Marcell but 42 percent of the total deposition. During hot, stagnant conditions, Crystal River heat input rates sometimes affected mercury concentrations and deposition at the Chasshowitzka station.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

After mercury has been deposited from the atmosphere, biological processes typically convert it to methylmercury. Oral exposure to methylmercury is the primary pathway to mercury exposure for humans. Mercury exposure has been associated with several health effects, including decreased performance on neurobehavioral tests in children and an increased risk of coronary disease in adults. As a result of these findings, the U.S. Environmental Protection Agency (EPA) promulgated the Clean Air Mercury Rule (CAMR) in May 2005 to regulate the mercury emissions from coal- and oil-fired utilities.[1]

As the EPA works to implement emissions standards for the toxic pollutant mercury and its compounds, investigations continue to determine the fate of mercury in the environment. Mercury is emitted from coal-fired power plants in elemental, oxidized, and particulate-bound states.[2] Wet and dry deposition processes may remove the emitted mercury from the atmosphere near the power plant, or the mercury may travel significant distances before being deposited.

Many studies described the transport, chemistry, and deposition of mercury in the atmosphere, on both local and global scales.[3-13] Weiss-Penzias *et al.*[13] examined the oxidation of gaseous elemental mercury in plumes in the marine boundary layer. Dvonch *et al.*[7] studied the effect of local anthropogenic sources on atmospheric deposition of mercury in south Florida. The modeling efforts of Cohen and others[5-6] examined the deposition of current and future mercury emissions to the Great Lakes, implicating sources as far away as Florida. Bullock's modeling assessment[3] concluded that Canada's emission sources do not significantly affect wet deposition rates of mercury over the United States. Tsiros' modeling assessment[12] found that his screening-level model compared well with field data and showed the effects of solar radiation, soil moisture, temperature, and wind conditions on the reduction of divalent mercury. Seigneur *et al.*[11] examined a model's performance to simulate the global cycling of atmospheric mercury and the mercury deposition to potentially sensitive areas; they found many consistencies with available measurements but continued to be limited by the uncertainties in emissions estimates, chemistry, and deposition processes.

The current study by the Maryland Power Plant Research Program (PPRP) aimed to identify whether or not a measurable quantity of the mercury is deposited near sources under any meteorological conditions. To examine large data sets or databases with many fields, the term knowledge discovery through databases (KDD) has been coined to describe this area of high performance computing. The application of KDD has been described as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data," and data mining is just one component of KDD that is defined as "extraction of patterns or models from observed data."[14]

The KDD techniques such as neural networks and decision trees are already familiar to investigators in the air quality field.[15-24] For example, McKendry[24] examined the performance of neural networks for forecasting particulate matter concentrations based on meteorological conditions. Chelani *et al.*[23] used neural networks to predict ambient concentrations of respirable particulate matter and toxic metals, and Dimopoulos et al.[25] combined neural network models to predict the spatial distribution of lead around a point emission source.

Neural networks have been useful for future predictions, but the rule induction approach offers descriptive models with more explanatory power. Rule induction techniques report that a list of attributes within the data set will predict a value within a particular confidence level and significance.[6] Through the use of rule induction techniques, it is sometimes possible to discern the conditions that lead to a relationship within computer-determined subsets of the data.

Mercury emissions are highly variable from individual coal-fired boilers and are dependent on chlorine content of the fuel, flue gas temperatures, boiler design, control device operation, ash characteristics, and other operating characteristics.[26-28] However, long-term measurements of mercury emissions (and their speciation) at power plants are not yet available. In this study, the identification of significant trends is performed by comparing ambient mercury measurements of wet deposition to the heat input rates at nearby power plants. Instead of simple regression analysis techniques, the authors employed a piecewise regression technique. The piecewise regression analysis (classifying the data into different subsets) will yield different rules (regression fits) for different meteorological conditions, unlike a simple regression analysis.

2. DATA PREPARATION

2.1 Site Selection

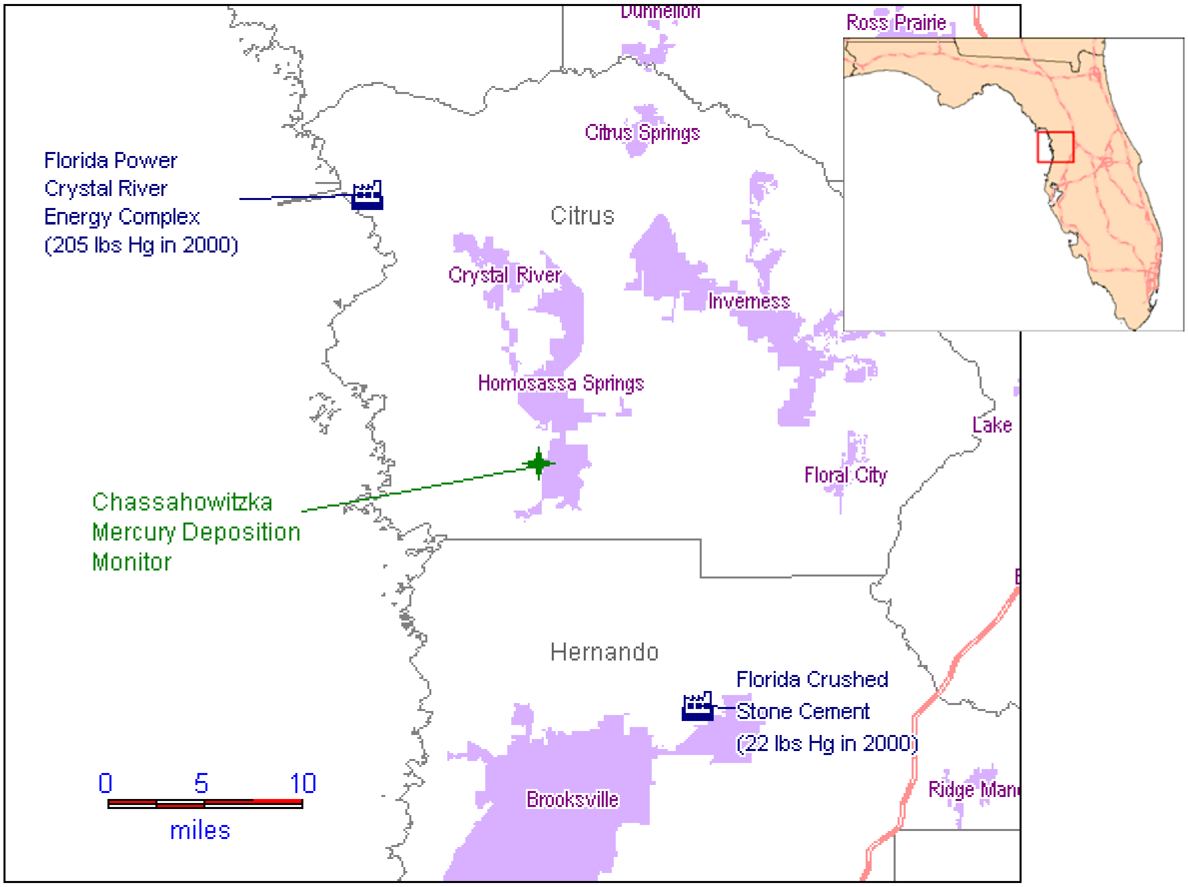
To understand how a single power plant might affect the mercury deposition readings at a monitor, it was important to choose ambient measurement sites that were not in close proximity to more than one major emissions source. Compliance with this goal does not imply that emissions from other large sources will not affect the mercury deposition rates, but only that short-term variations in the rates will be driven by the local source. Even remote areas in the eastern United States show significant mercury deposition.

Only two eastern sites were identified in the National Acid Deposition Program's Mercury Deposition Network (NADP/MDN) that had analyzed at least five years of data and were located far from more than a single large mercury emissions source. The first is the Chassahowitzka National Wildlife Refuge Station located in Citrus County, Florida at an elevation of ten feet. Since July 1997, the U.S. Fish and Wildlife Service-Chassahowitzka National Wildlife Refuge has operated this station, shown in Figure 1.[29]

Figure 1. Mercury Deposition Network site at Chassahowitzka National Wildlife Refuge

Figure 2 shows the location of the Chassahowitzka station and the nearest mercury emissions sources, as reported in the EPA's 2000 Toxics Release Inventory. Florida Power's Crystal River Energy Complex is located 16 miles northwest of the Chassahowitzka station and reported 205 pounds of mercury emitted in 2000. The next nearest emissions source (Florida Crushed Stone Cement) was located to the southeast but only emitted 22 pounds of mercury in 2000. After that, the nearest mercury emissions sources were almost 50 miles away from Chassahowitzka.

Figure 2. Mercury emissions sources near Chassahowitzka mercury deposition monitor



The Crystal River Energy Complex operates four units with similar heat input capacities. The data from Units 1, 2, 4, and 5 covered the first quarter in 1997 through the fourth quarter in 2003. A description of the units appears in Table 1.

Table 1. Units at the Crystal River Energy Complex

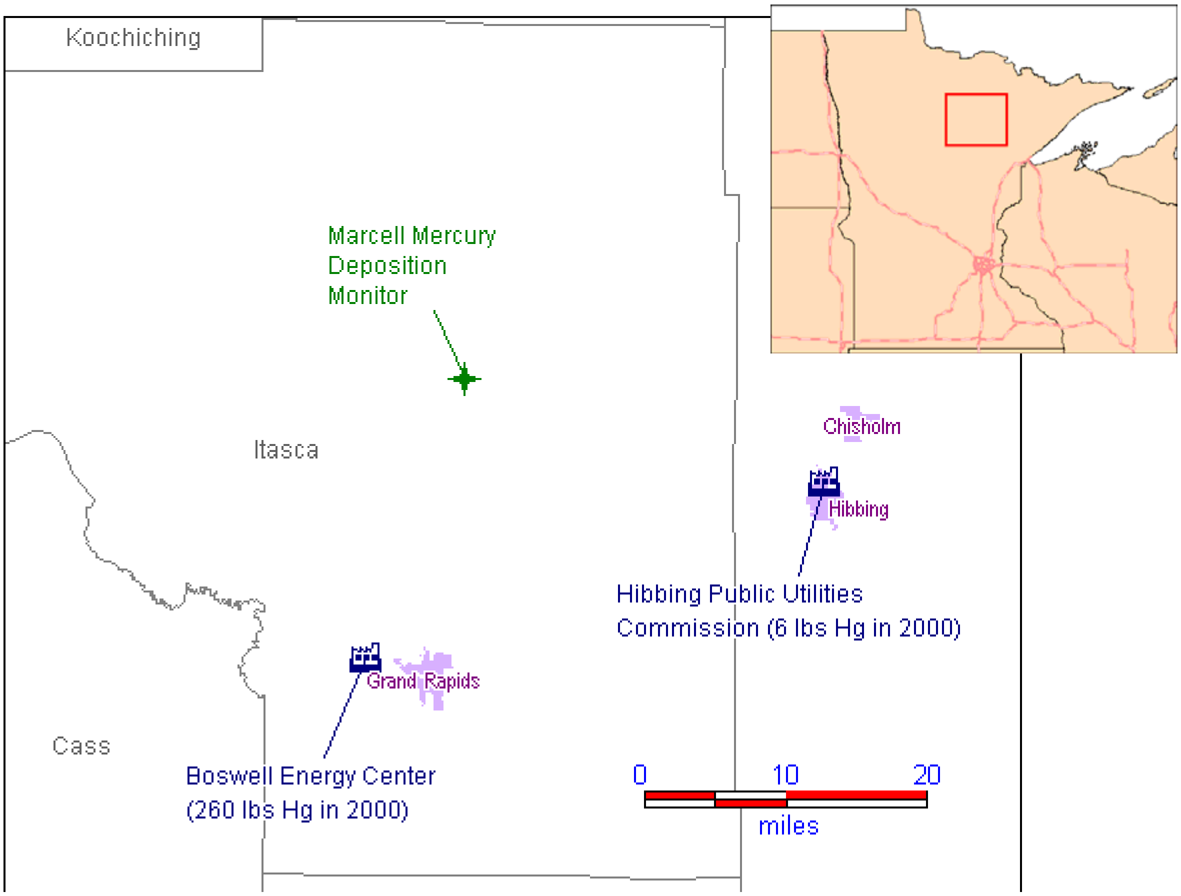
Unit ID	Boiler Description	Boiler Capacity (mmBtu/hr)	Controls Description
1	Tangentially fired	3,750	Low NO_x burners and electrostatic precipitator
2	Tangentially fired	4,795	
4	Dry-bottom wall-fired	6,665	
5	Dry-bottom wall-fired	6,665	

The second site was the Marcell Experimental Forest Station located in Itasca County, Minnesota at an elevation of 1,400 feet. Since February 1995 the U.S. Forest Service-North Central Research Station has operated this station, shown in Figure 3.[29]

Figure 3. Mercury Deposition Network site at Marcell Experimental Forest Station

Figure 4 shows the location of the Marcell station and the nearest mercury emissions sources, as reported in the 2000 Toxics Release Inventory. The Boswell Energy Center is located 21 miles south of the Marcell station and reported 260 pounds of mercury emitted in 2000. The next nearest emissions source (Hibbing Public Utilities Commission) was located 26 miles to the east but only emitted 6 pounds of mercury in 2000. After that, the nearest mercury emissions sources were over 60 miles away from the Marcell station.

Figure 4. Mercury emissions sources near Marcell mercury deposition monitor



The Boswell Energy Center operates four units with different heat input capacities. The data for these units covered 1995 through 2003. A description of the units appears in Table 2.

Table 2. Units at the Boswell Energy Center.

Unit ID	Boiler Description	Boiler Capacity (mmBtu/hr)	Controls Description
1	Dry-bottom wall-fired	750	Baghouse
2	Dry-bottom wall-fired	750	Baghouse
3	Tangentially fired	3,355	Wet scrubber
4	Tangentially fired	5,109	Low NO_x burner with closed coupled overfire air, wet scrubber, and electrostatic precipitator on the reheat bypass

2.2 Data Collection

Mercury data were retrieved for the Chassohowitzka and Marcell sites for the time period between July 1997 to December 2002.[29] The mercury data consisted of the site name, sample collection start time, sample collection end time, amount of precipitation in the rain gauge, mercury concentration, mercury deposition (calculated), a quality rating note, and the sample type. Mercury concentration (Hg Concn) was directly measured at the sites, and each mercury weekly deposition value was calculated by multiplying the concentration by the amount of precipitation. The quality rating can be an A (172 samples), B (349 samples), or C (11 samples). An "A" rating means that the data was fully qualified with no problems, and "B" ratings indicated valid data with minor problems that are generally only used for summary statistics. A "C" rating indicates that the data was invalid and it is not used in any summary statistics; data with "C" ratings were excluded from these analyses. The sample type indicated whether measurable precipitation, low precipitation, or trace precipitation was measured in the sample bottle catch. Precipitation data from the NADP network[29] was used to determine the hours with precipitation at the mercury sites.

Hourly surface meteorology data were collected from the National Climatic Data Center for sites that were near the mercury site with similar topography.[30] Temperature, relative humidity, wind speed, and wind direction were considered. The two sites chosen were Brooksville, Florida and Park Rapids, Minnesota. The local climatological data with unedited hourly observations were collected for the period from June 1996 to December 2002 for the Brooksville site and from November 1996 to December 2002 for Park Rapids.

Hourly plant emission data (heat input, sulfur dioxide emissions, and carbon dioxide emissions) were retrieved for the Clay Boswell plant in Minnesota and for the Crystal River plant in Florida, the major mercury sources near the monitor sites. EGRID data were retrieved from the EPA's Clean Energy site.[31] The heat input data for each hour were summed for all operating units. The hourly meteorology and emissions data were tallied and averaged to match the weekly mercury data.

An alternate approach would have been to estimate the mercury emissions from heat input measurements for the individual boilers. To do this, the heat input would be multiplied by the mercury emissions factor for the unit and then summed to calculate the total mercury emissions. However, the mercury testing represented only a week of collected data, and reports indicate that mercury emission rates are highly variable from a single unit, changing more than an order of magnitude with changes in coal blends.[28,35] These uncertainties suggested that a direct comparison with heat input would be easier to substantiate than estimated mercury emissions.

2.3 Data Mining

For each mercury deposition site, a single data set was arranged by date for input into RuleQuest's Cubist 1.11 software. The Cubist program constructs an unconventional type of regression tree, with the leaves containing linear models instead of discrete values. A classification tree would categorize the predictions into discrete subsets, but the regression tree predicts actual continuous values. Cubist reads and initially groups the training cases. Then a production rules approach is followed to simplify the decision trees and create comprehensible rules: the more restrictive conditions are evaluated and removed through Fisher's exact test.[32-34] After this process, the final rules are formed by evaluating the effects of all rules simultaneously, and the training and test data are evaluated. The final rules from Cubist often overlap so that multiple rules describe individual data instances.

Unlike neural networks, the Cubist program generates a model with rules that describe the relationships between the independent and dependent parameters in the data set. The models for these small data sets required less than one second to complete. A typical rule to predict mercury deposition (in ng/m^2) reads:

If

Weekly Precipitation $\leq$ 19.4 mm

Then

Chassahowitzka Mercury Deposition = -71.2 + 10.9 × (Precipitation in mm)
+ 858 × (Frequency of Variable Winds as a percentage)
+ 0.006 × (Crystal River Heat Input Rate as mmBtu/hr)
- 11 × (Average Wind Speed in knots)

[98 cases, mean 99.7 ng/m^2, range 2.9 to 471 ng/m^2, estimated error 48 ng/m^2]

The "if" portion represents the conditions of the rule, and the "then" portion shows the linear model. The variables appearing in the "If" statement will be referred to as the Conditional Parameters (e.g., precipitation in the sample rule), and the variables appearing in the "Then" statement indicate the Equation Parameters.

Since Cubist operates via piecewise regression (each rule being applicable under certain conditions), it can be more powerful than a multivariate linear model because it allows variables to be weighted differently as conditions change. When rules overlap, the predicted values from both models are averaged. When the number of parameters is small, Cubist can combine a rule-based model with a nearest-neighbor model to improve the predictive accuracy. When the Cubist program determined that a significant improvement could be made by including nearest-neighbor methodology, the "Rules and Instances" option was used to predict the deposition more accurately. The minimum rule cover represents the number of cases that must be included in each conditional subset in order to build a rule. The minimum rule cover was set to two percent for these model sets. The model rules were also allowed to extrapolate ten percent beyond the range for the target values set by a rule.

Using the piecewise regression approach, several models with different detailed rules can yield the same degree of accuracy. Therefore, the individual rules in one particular exemplar model should be viewed as merely indicative of representative dependencies between the variables.

Since heating and cooling needs drive demand on the power plants, ambient temperatures are sometimes correlated with power plant heat input rates.[27] This interrelationship can confound the determination of the correlations with a third parameter (e.g., ambient concentrations). Therefore, ambient temperature was removed from the rule induction process as an equation parameter to allow the heat input dependence to be recognized. However, ambient temperature (considered in 5-degree increments) remained an important conditional parameter. In a test run, temperature was included as an equation parameter in the rule induction process, but the errors were not substantially reduced.

3. RESULTS AND DISCUSSION

3.1 Presentation of Results

Many models could be developed to describe the data, but many of these models show similar fits to the data. Both mercury concentration and deposition were chosen as the target parameters for the Cubist models, and some models even targeted the utility heat input rates based on the mercury levels (not presented). The Cubist model runs were varied to alter the target parameters, the meteorological data set to use, and whether or not to include outliers. Tables 3 and 4 present brief summaries of the significant runs and identify the important parameters.

The runs with a full week of meteorology considered the meteorological averages (wind speed, wind direction, temperature, and relative humidity) over the course of the whole week, and not just during the precipitation period. The mercury concentrations/depositions from the data sets had four extreme outliers (more than three standard deviations from the average), and these outliers affected the derived models in runs where they were included. The average errors (average of the absolute values of the biases), relative errors, and correlation coefficients were also determined for the data sets. The relative error is calculated as the absolute value of the ratio of the average model error to the error that would have been measured if the mean value had been predicted each time. The correlation coefficients describe the best linear fit between the observations and the predictions.

If the Cubist model determined that the model performance could be improved by using a nearest neighbor approach to modify the output, then the number of nearest neighbors is reported in the tables. However, Cubist found that models for only three of the data sets improved by use of nearest neighbor techniques, likely because the data sets were small. The numbers of conditional rules used to describe the model are also reported in Tables 3 and 4.

When the Heat Input Rate from Crystal River or Boswell appeared in the model equations, the maximum and average changes from heat input were calculated. The maximum change was calculated by multiplying the highest heat input rate (based on the complete data set for that power plant) by the model's coefficient before the heat input parameter. This number was then corrected for the frequency by multiplying it by the number of cases in that subset of data and dividing by the total number of cases. For example, in the third run for Chassahowitzka (Table 3), one equation had a coefficient of 7e-5 ng-hr/l-mmBtu, the equation was valid in 46 of the 199 cases, and the Crystal River power plant data had a maximum heat input rate of 25,300 mmBtu/hr. Therefore, the maximum change was calculated as

$$\text{Maximum Change from Heat Input} = 7\text{e-}5 \times \frac{46}{199} \times 25{,}300 = 0.41 \text{ ng/l}$$

The Average Change from Heat Input was calculated in a similar fashion, but the heat input rate was taken as the average of the subset that met the conditions for the rule.

The remaining rows in Tables 3 and 4 show which parameters appeared in the Cubist models. The wind direction frequencies (e.g., frequency of wind from the northeast) appear in most of the models, but their interdependencies (i.e., they must add to one hundred percent) make their individual influences difficult to evaluate. The wide divergence in the parameters used in the models shows that the models are not unique in their fits to the data, so significant confidence should not be placed in any single model. Instead, an evaluation of the models and the model parameters should lead investigators to identify the important information "nuggets."

Table 3. Summary of significant runs for Chassahowitzka station (median concentration = 12.3 ng/l and median weekly deposition = 242 ng/m^2).

Target Parameter	**Hg Concn**	**Hg Concn**	**Hg Concn**	**Hg Deposition**	**Hg Deposition**	**Hg Deposition**
Full week of meteorology?	No	Yes	Yes	No	Yes	Yes
Limit to 3 standard deviations?	No	Yes	No	No	Yes	No
Number of Cases	198	196	199	198	196	199
Average Error	4.8	5.4	5.3	81	73	95
Relative Error	0.64	0.76	0.67	0.25	0.22	0.29
Correlation Coefficient	0.75	0.58	0.75	0.96	0.96	0.94
Number of Nearest Neighbors	9	Instances not used	Instances not used	9	9	Instances not used
Number of Rules	8	3	6	13	15	6
Maximum Change from Heat Input	27 ng/m^3	--	0.41 ng/m^3	--	52 ng/m^2	126 ng/m^2
Average Change from Heat Input	25 ng/m^3	--	0.33 ng/m^3	--	44 ng/m^2	94 ng/m^2
Conditional Parameters (not including directional wind frequencies)						
Precipitation	X	X	X	X	X	X
Wind Speed			X	X	X	X
Temperature	X	X	X		X	X
Month				X		
Relative Humidity				X		
Calm Wind Frequency	X			X		
Crystal River Heat Input Rate	X					
Equation Parameters (not including directional wind frequencies)						
Precipitation	X	X	X	X	X	X
Wind Speed	X	X	X	X	X	X
Year	X	X				
Relative Humidity	X		X	X		
Crystal River Heat Input Rate	X		X		X	X

Table 4. Summary of significant runs for Marcell station
(median concentration = 9.0 ng/l and median weekly deposition = 83 ng/m^2).

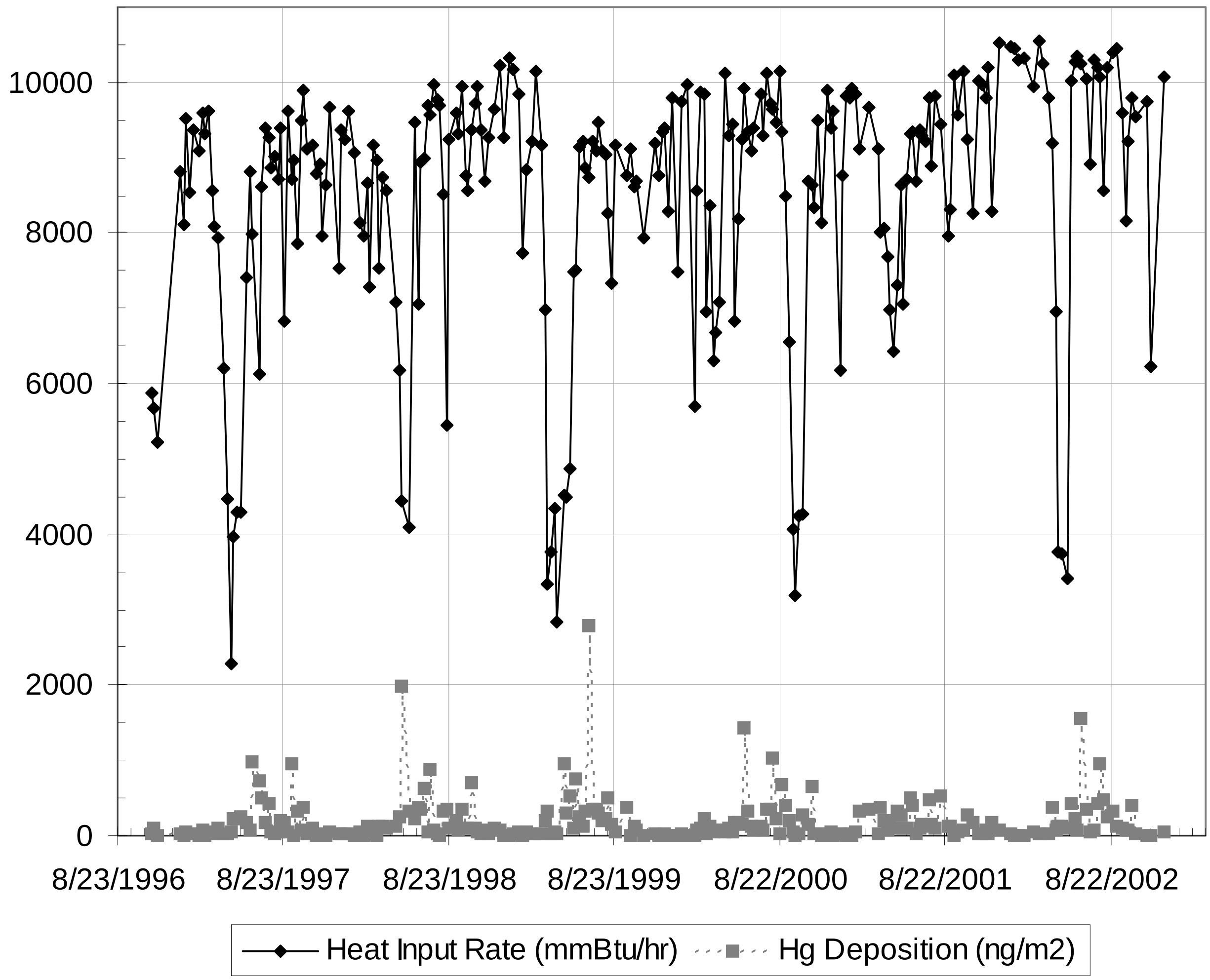
Target Parameter	Hg Concn	Hg Concn	Hg Concn	Hg Concn	Hg Deposition	Hg Deposition	Hg Deposition
Full week of meteorology?	No	No	Yes	Yes	No	Yes	Yes
Limit to 3 standard deviations?	No	Yes	Yes	No	No	Yes	No
Number of Cases	250	249	249	250	250	249	250
Average Error	7.6	5.5	5.7	9.3	75	61	62
Relative Error	0.75	0.67	0.71	0.93	0.41	0.33	0.34
Correlation Coefficient	0.34	0.69	0.71	0.16	0.89	0.93	0.92
Number of Nearest Neighbors	Instances not used	Instances not used	Instances not used	Instances not used	Instances not used	Instances not used	Instances not used
Number of Rules	3	6	5	1	5	8	9
Maximum Change from Heat Input	--	--	--	--	--	-60 ng/m^2	-29 ng/m^2
Average Change from Heat Input	--	--	--	--	--	-50 ng/m^2	-23 ng/m^2
Conditional Parameters (not including directional wind frequencies)							
Precipitation	X	X	X		X	X	X
Wind Speed							X
Temperature	X	X	X		X	X	X
Relative Humidity			X		X	X	X
Boswell Heat Input Rate		X					
Equation Parameters (not including directional wind frequencies)							
Precipitation		X			X	X	X
Wind Speed		X	X			X	X
Relative Humidity	X	X	X	X	X	X	X
Calm Wind Frequency		X					
Boswell Heat Input Rate						X	X

3.2 Analysis of Marcell Results

Table 4 shows a surprising dependence in the last two runs. A negative dependence appears in the comparison between the Boswell heat input and the local mercury deposition. The conditional dependences for the equations with the negative dependence specify that the precipitation be less than or equal to 25.1 mm and that the temperature be over 10°C (69 cases), and a second condition (for the sixth run) that the precipitation be greater than 25.1 mm and the temperature be less than or equal to 15°C (24 cases). For 20 of the 21 cases that meet the first condition and have heat input rates below 8500 mmBtu/hr, the mercury deposition was greater than 100 ng/m^2. However, the median for all 250 cases in the entire data set is just 83 ng/m^2. Therefore, these low-precipitation, high-temperature cases indicate that increases in ambient mercury deposition may be occurring when the Boswell Energy Center is operating well below capacity.

Figure 5 indicates that most spikes in mercury deposition over 100 ng/m^2 occur after the Boswell Energy Center has operated at a reduced capacity. Further investigation showed that most of the spikes over 900 ng/m^2 were associated with the shutdown or startup of Boilers #1 or 4; other spikes occurred during weeks of constant rain with winds from the south. Therefore, the rises in mercury deposition may be associated with startup/shutdown conditions for at least two of the boilers. Additional investigation would be necessary to confirm this behavior and may reveal why the influence was not seen in the runs that used mercury concentration as the target parameter.

Figure 5. Time series comparison of Boswell Energy Center heat input rates and Marcell mercury deposition



The spikes in ambient mercury deposition during startup and shutdown may be related to the mercury speciation at the outlet. The DOE Emissions Test Report for Boswell Energy Center[35] listed the compositions shown in Table 5. Because elemental mercury disperses in the atmosphere more readily than divalent mercury, mercury in the divalent form will deposit more readily and be detected by the nearby wet deposition monitor. If mercury speciation from Boswell Energy Center has more divalent mercury than elemental mercury during startup phases, the mercury deposition may spike during these periods.

Table 5. Mercury speciation at Boswell Energy Center (unit outlet).[35]

Mercury Species	Unit 2	Unit 3	Unit 4
Percent elemental mercury	22	99	91
Percent oxidized mercury (primarily divalent mercury)	76	<1	6
Percent particle-bound mercury	<2	<0.1	3

The mercury speciation can also change in the control devices.[26] If the temperature in an electrostatic precipitator is not too high (less than 140°C), mercuric chloride ($HgCl_2$) is less volatile and can condense on fly-ash particles. When hydrogen chloride (HCl) concentrations are high, more mercury is collected in the fly ash.

It is also possible that the mercury emissions during the startup phase of a unit are not at a steady state. Meij, Vredenbregt, and Winkel[26] describe a non-steady state process when more mercury appears to leave a flue gas desulfurization system than enter it. Meij *et al.* attributed the difference to the previous adsorption of mercury onto the scrubber walls, such that the equilibrium changes and mercury is released when the flue gases change flow rate, temperature, or mercury and chloride concentrations.

The significance of the deposition during these spikes should not be underestimated. The spikes above 500 ng/m^2 represent only eight percent of the weekly data at Marcell but 42 percent of the total deposition. The largest spike occurred during the week of June 29, 1999 (2800 ng/m^2), and the deposition during this one week is equivalent to the total deposition over the following 23 weeks (7.5 months).

During the startup at the Boswell Energy Center, it should first be determined which units and control devices were starting up when the spikes were observed. Further investigations should then be able to determine if the total mercury releases were increasing or if more mercury was oxidized in those specific units and control devices during the startup.

3.3 Analysis of Chassahowitzka Results

In the fifth and sixth runs for the Chassahowitzka station (Table 3), one common condition related the mercury deposition directly to the heat input rate at Crystal River Energy Complex. In 15 cases (13 of which occurred in the summer), the following rule applied:

If

Precipitation > 19.4 mm
Average wind speed ≤ 4.65 knots
Wind frequency from northeast ≤ 11.9 percent

Then

$$\text{Mercury deposition} = -709 + 2490 \times (\text{wind frequency from northeast}) + 10.6 \times (\text{precipitation}) + 0.027 \times (\text{Crystal River heat input rate})$$

The cases did not simultaneously obey another rule. The average wind frequency from the northeast was just six percent in these cases, and 46 percent of the data showed either calm or variable winds (with 40 percent from the southeast and southwest). This rule indicates that the Crystal River operation affects the mercury deposition under low wind speeds when the winds are generally calm. Under mostly stagnant conditions, the local air mass is likely well-mixed and a particular surface wind direction plays only a minor role.

Similarly the first run from the Chassahowitzka station indicates that the Crystal River operation affects the mercury concentration, but this strong effect was not observed in the second and third runs. The difference between these runs is that the first run considers only meteorological conditions and heat inputs during the precipitation event, suggesting that the weekly averages are not as useful for discerning the influences of the Crystal River operation. The errors were slightly smaller for the first run than the other two runs.

The first run contains the Crystal River heat input rate as a conditional or equation parameter in seven of its eight rules, covering 174 of the 198 cases. The Average Change from Heat Input is 25 ng/l and is influenced by five rules. However, one rule is responsible for 86 percent of the Average Change from Heat Input and only applies in five cases:

If Frequency of calm winds > 28 %
Crystal River heat input rate > 23,700 mmBtu/hr
Then
Mercury concentration = –875 + 0.038 × (Crystal River heat input rate)

The mercury concentrations in these five cases ranged from 26 to 61 ng/l, and the meteorological conditions suggested average temperatures over 25°C, mean relative humidities between 83 and 90 percent, and calm winds during the precipitation event. The average wind speeds were less than 3.6 knots, and the calm and variable winds accounted for 46 percent of the wind directions. The meteorological and heat input data during precipitation events in those five weeks were compiled from 24 to 108 hourly measurements.

Dvonch *et al.* found that local anthropogenic sources played a significant role in the local atmospheric deposition of mercury in south Florida during summer months.[7] Dvonch *et al.* found that the wet deposition from local emissions sources was associated with convective precipitation events that result from storm cells much smaller in size and duration than frontal systems. The isolated convective storms are common in south Florida during summer months.

Similar meteorological conditions may lead to the influence of Crystal River on the Chassahowitzka site. Further investigation might reveal the nature of the five storm events (occurring in July, August, and September) that Cubist identified. If certain meteorological events are associated with local mercury deposition, additional mercury controls might be useful on local emissions sources during the time of year when these types of storm events are expected.

4. CONCLUSIONS

The goal of this study was to determine if local emissions sources affect nearby ambient mercury deposition. By comparing the meteorological conditions and power plant heat input rates with the weekly mercury deposition, the Cubist models helped identify conditions where local emissions sources influenced nearby monitors. The four most significant findings were:

- Mercury deposition measurements spiked at the Marcell station when the Clay Boswell heat input rates increased. This finding is likely connected to increased mercury emissions during unit startups or shutdowns.
- The spikes in mercury deposition represent a significant portion of the annual deposition at the Marcell monitor site. Reductions in the spikes could significantly lower the annual mercury deposition.
- The mercury deposition values during the summer at the Chassahowitzka site were influenced by emissions at Crystal River Energy Complex under low wind speeds. This finding was reflected by weekly average data for the power plant and meteorology.

- Similarly, in five different weeks, the summer mercury concentrations at the Chassahowitzka site were influenced by emissions at Crystal River Energy Complex under low wind speeds and high temperatures. This finding was reflected by data for the power plant and meteorology that were collected solely during the precipitation event hours.

These findings suggest that conditions exist when these two large emissions sources influence the local mercury deposition. Because the climatology of northern Minnesota and Florida are considerably different, it is not surprising that the data mining exercise identified different trends (one based on heat input changes and the second on wind speeds and temperature).

This demonstration illustrates the utility of data mining tools, even with small data sets. One of the advantages of the data mining tools is that the data subsets can be determined where important trends can be found. In this study, the rule induction tools helped identify interesting cases that shared a characteristic (i.e., dependence on power plant emissions) and should be examined with greater scrutiny.

5. REFERENCES

1. Standards of Performance for New and Existing Stationary Sources: Electric Utility Steam Generating Units. *Fed. Regist.*, **2005**, 70, May 18, 2005.
2. U.S. Department of Energy. Calculation of Possible Mercury MACT Floor Values for Coal-Fired Utilities: Influence of Variability and Approach. December 2003.
3. Bullock, O. Russell, Jr. Modeling assessment of transport and deposition patterns of anthropogenic mercury air emissions in the United States and Canada; *The Science of the Total Environment*. **2000**, 259, 145-157.
4. Bullock, O. Russell, Jr.; Brehme, Katherine A. Atmospheric mercury simulation using the CMAQ model: formulation description and analysis of wet deposition results; *Atmospheric Environment*. **2002**, 36, 2135-2146.
5. Cohen, Mark D.; Miller, Paul J.; Atmospheric Mercury Deposition Impacts of Future Electric Power Generation; *Commission for Environmental Cooperation*. **2003**, 1-11.
6. Cohen, Mark D.; Artz, Richard; Draxler, Roland; Miller, Paul; Poissant, Laurier; Niemi, David; Ratte, Dominique; Deslauriers, Marc; Duval, Roch; Laurin, Rachelle; Slotnick, Jennifer; Nettesheim, Todd; McDonald, John. Modeling the atmospheric transport and deposition of mercury to the Great Lakes; *Environmental Research*, accepted for publication.
7. Dvonch, J.T.; Graney, J.R.; Marsik, F.J.; Keeler, G.J.; Stevens, R.K. An investigation of source-receptor relationships for mercury in south Florida using event precipitation data; *The Science of the Total Environment*. **1998**, 213, 95-108.
8. Lin, Che-Jen; Pehkonen, Simo O. The chemistry of atmospheric mercury: a review; *Atmospheric Environment*. **1999**, 33, 2067-2079.
9. Pai, Prasad; Karamchandani, Prakash; Seigneur, Christian. Simulation of the Regional Atmospheric Transport and Fate of Mercury using a Comprehensive Eulerian Model; *Atmospheric Environment*. **1997**, 31, 2717-2732.
10. Ryaboshapko, Alexey; Bullock, Russell; Ebinghaus, Ralf; Ilyin, Ilia; Lohman, Kristen; Munthe, John; Petersen, Gerhard; Seigneur, Christian; Wangberg, Ingvar. Comparison of mercury chemistry models; *Atmospheric Environment*. **2002**, 36, 3881-3898.

11. Seigneur, Christian; Karamchandani, Prakash; Lohman, Kristen; Vijayaraghavan, Krish; Shia, Run-Lie. Multiscale modeling of the atmospheric fate and transport of mercury; *J. of Geophysical Research*. **2001**, 106, 27, 795-27, 809.
12. Tsiros, Ioannis. Modeling Assessment of Air Emission Flux of Mercury from Soils in Terrestrial Landscape Components: Model Tests and Sensitivities; *J. Air & Waste Manage. Assoc*. **2002**, 52, 339-348.
13. Weiss-Penzias, Peter; Jaffe, Daniel A.; McClintick, Anna; Prestbo, Eric M.; Landis, Matthew S. Gaseous Elemental Mercury in the Marine Boundary Layer: Evidence for Rapid Removal in Anthropogenic Pollution; *Environ. Sci. Technol.* **2003**, 37, 3755-3763.
14. Goebel, M.; Gruenwald, L. A Survey of Data Mining and Knowledge Discovery Software Tools; *SIGKDD Explorations: Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining*. **1999**, *1*, 20-33.
15. Wienke, D.; Gao, N.; Hopke, P.K. Multiple Site Receptor Modeling with a Minimal Spanning Tree Combined with a Neural Network; *Environ. Sci. Technol.* **1994**, *28*, 1023-1030.
16. Comrie, A.C. Comparing Neural Networks and Regression Models for Ozone Forecasting; *J. Air & Waste Manage. Assoc.* **1997**, *47*, 653-663.
17. Guardani, R.; Nascimento, C.A.O.; Guardani, M.L.G.; Martins, M.H.R.B.; Romano, J. Study of Atmospheric Ozone Formation by Means of a Neural Network-Based Model; *J. Air & Waste Manage. Assoc.* **1999**, *49*, 316-323.
18. Kao, J.-J.; Huang, S.-S. Forecasts Using Neural Network versus Box-Jenkins Methodology for Ambient Air Quality Monitoring Data; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 219-226.
19. Melas, D.; Kioutsioukis, I.; Ziomas, I.C. Neural Network Model for Predicting Peak Photochemical Pollutant Levels; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 495-501.
20. Hadjiiski, L.; Hopke, P. Application of Artificial Neural Networks to Modeling and Prediction of Ambient Ozone Concentrations; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 894-901.
21. Cobourn, W.G.; Dolcine, L.; French, M.; Hubbard, M.C. A Comparison of Nonlinear Regression and Neural Network models for Ground-Level Ozone Forecasting; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 1999-2009.
22. Gardner, M.; Dorling, S. Artificial Neural Network-Derived Trends in Daily Maximum Surface Ozone Concentrations; *J. Air & Waste Manage. Assoc.* **2001**, *51*, 1202-1210.
23. Chelani, A.B.; Gajghate, D.G.; Hasan, M.Z. Prediction of Ambient PM_{10} and Toxic Metals Using Artificial Neural Networks; *J. Air & Waste Manage. Assoc.* **2002**, *52*, 805-810.
24. McKendry, I.G. Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM_{10} and $PM_{2.5}$) Forecasting; *J. Air & Waste Manage. Assoc.* **2002**, *52*, 1096-1101.
25. Dimopoulos, Ioannis F.; Tsiros, Ioannis X.; Serelis, Konstantinos; Chronopoulou, Aikaterini. Combining Neural Network Models to Predict Spatial Patterns of Airborne Pollutant Accumulation in Soils around an Industrial Point Emission Source; *J. Air & Waste Manage. Assoc*. **2004**, 54, 1506-1515.
26. Meij, Ruud; Vredenbregt Leo H.J.; Winkel, Henk te. The Fate and Behavior of Mercury in Coal-Fired Power Plants; *J. Air & Waste Manage. Assoc*. **2002**, 52, 912-917.
27. PPRP-ERM-SAIC Meeting. Pilot Study: Data Mining to Determine Local Effects of Mercury Emissions. November 29, 2004.

28. Emission Strategies, Inc. Mercury Measurement and Characterization at Constellation Power Source Generation's Coal-Fired Units. May 2004.
29. National Atmospheric Deposition Program. Sites in the NADP/MDN Network. http://nadp.sws.uiuc.edu/nadpdata/mdnsites.asp (accessed January 2004).
30. National Climatic Data Center. Unedited Local Climatological Data. http://cdo.ncdc.noaa.gov/ulcd/ULCD (accessed January 2004).
31. U.S. Environmental Protection Agency. Clean Energy – eGrid. http://www.epa.gov/cleanenergy/egrid/index.htm (accessed January 2004).
32. Quinlan, J.R. Introduction. *C4.5: Programs for Machine Learning*; The Morgan Kaufmann Series in Machine Learning; Morgan Kaufmann Publishers: San Mateo, California, 1993; 5-12.
33. Quinlan, J.R. Simplifying decision trees. *International Journal of Human-Computer Studies*. **1999**, *51*, 497-510.
34. Personal communication with Ross Quinlan, RuleQuest Research. October 8, 2002.
35. Roy F. Weston, Inc. DOE Project: Assessment of Speciated Mercury Emissions from Three Coal Fired Boilers. May 2000.