FUZZY LOGIC MODELING OF SURFACE OZONE CONCENTRATIONS

Rachel Mintz
June 3, 2005
OVERVIEW

- Fuzzy logic
- Modified Learning from Examples
- Edmonton Predictive Model
- Edmonton Forecast Model
- NEW! Edmonton PM2.5 Model
AIR POLLUTION MODELS

Deterministic models
- EPA’s UAM model
- Environment Canada’s CHRONOS model

Empirical models
- Environment Canada’s CANFIS model
WHAT IS FUZZY LOGIC?

Values from 0 to 1
**FUZZY LOGIC: IF-THEN RULES**

**TWO INPUTS:**
- Wind Speed
- Wind Direction

**ONE OUTPUT:**
- Ozone Concentration

1. **If WIND SPEED is LOW** and **If WIND DIRECTION is NORTHERLY**
   - Then OZONE is MEDIUM

2. **If WIND SPEED is MEDIUM** and **If WIND DIRECTION is WESTERLY**
   - Then OZONE is LOW

---

**OZONE CONCENTRATION (ppm):**
- 0.002
- 0.02
- 0.039
FUZZY LOGIC: IF-THEN RULES

INPUTS:
- Wind speed 4.0 m/s and
- Wind Direction 290 Deg.

OUTPUTS:
- RED and BLUE Rules

DEFUZZIFICATION

\[ f(x) = \frac{\sum_{i=1}^{3} b_i \prod_{j=1}^{2} \mu_{ij}}{\sum_{i=1}^{3} \prod_{j=1}^{2} \mu_{ij}} \]

\( b = \text{OZONE CONCENTRATION (ppm)} \)
Why Use Fuzzy Logic in Air Pollution Modeling?

• Approximate solutions are acceptable.
• Variables involve measurements of uncertainty or observational errors.
• Input-output relationships exist but are not well-defined or even consistent.
• Mathematical formulas are either unknown or complex.
• Emissions inventory is often unknown.
• Objective: Create a rule base (membership functions) to describe the system.
• Experiential learning method.
• Steps:
  – Training data
  – Create rule base
  – Test data
First training data point:

- WSP = 3.0 m/s
- WDR = 285 deg.

Output:

- Ozone = 0.002 ppm

\[ b_1 = 0.002 \]
Next training data point:

WSP = 2.0 m/s
WDR = 325 deg.

Output:
Ozone = 0.02 ppm

If the difference between the fuzzy output and actual output exceed a user defined tolerance, then a new rule is added.
MLFE MODEL

Rule 1:
\[ b_1 = 0.002 \]

Rule 2:
\[ b_2 = 0.02 \]

The new spread is determined using either the max or min distance between centers.

\[ \sigma^{new} = \min \left| c^{new} - c^i \right| \]

or

\[ \sigma^{new} = \max \left| c^{new} - c^i \right| \]
EDMONTON PREDICTIVE MODEL

- 3 Months: July, August, September.
- Training Data: 2000 and Test Data: 2001
- Perfect prognosis: past meteorological events are used as inputs.
• Different training data sets tested.
• Input combinations tested with past ozone, temperature, relative humidity, wind direction, and wind speed.
• Three and four input MLFE models.
• Different initial spread specifications tested.
• Maximum and minimum distance criteria tested.
BEST MLFE MODEL

- City of Edmonton during the summer months.
- Four-Input: \{Past O_3, WSP, Temp, RH\}
- Maximum Distance Criteria
- Range specification as the initial spread.
INPUT VARIABLES and TUNING PARAMETERS

- Compare residual error:
  \[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{predicted} - y_{observed})^2} \]
  \[ MAE = \frac{1}{n} \sum_{i=1}^{n} |y_{predicted} - y_{observed}| \]

- Compare scatter plots of observed versus predicted:
  - Slope
  - Coefficient of Determination \((R^2)\)
Three-Input Variable Combination

RMSE and MAE (ppm)

- July RMSE
- July MAE
- August RMSE
- August MAE
- September RMSE
- September MAE

Three-Input Variable Combination

- RH Temp Past O3
- Temp WDR Past O3
- RH WDR Past O3
- RH WSP Past O3
- WDR WSP Past O3
- Temp WSP Past O3
- RH Temp WDR
- Temp WDR WSP
Three-Input Variable Combination

Scatter Plot Measurement

July Coefficient of Determination
July Slope
August Coefficient of Determination
August Slope
September Coefficient of Determination
September Slope

RH Temp Past O3
Temp WDR Past O3
RH WDR Past O3
RH WSP Past O3
WDR WSP Past O3
Temp WSP Past O3
RH Temp WDR
Temp WDR WSP
Two different sets of training data

- July A RMSE
- July A MAE
- July B RMSE
- July B MAE

Input Combinations

- Four-input variable combinations
  - Temp, RH
  - Past O3
  - WDR, Temp
  - Past O3, WSP
- Three-input variable combination
  - Temp, RH
  - Past O3, WSP

July B produces a model with better performance. July B is only slightly better.

Three different training data sets were used. Two different sets of training data were used for the three-input variable combination.
• Objective: predict highest ozone concentration.
• Add “expert” training data points to clearly specify the highest ozone concentrations.
• Re-construct the rule base.
• 6 training data points are added: Temp high (>26C), WSP is low (<6km/h) and RH is low (<40%).
• Corresponding outputs for expert training data points are set at ozone levels >0.08ppm.
<table>
<thead>
<tr>
<th></th>
<th>July</th>
<th>August</th>
<th>September</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLFE</td>
<td>CHRONOS</td>
<td>CANFIS</td>
<td>MLFE</td>
</tr>
<tr>
<td>RMSE (ppb)</td>
<td>5.7</td>
<td>11.6</td>
<td>10.2</td>
<td>7.5</td>
</tr>
<tr>
<td>MAE (ppb)</td>
<td>4.8</td>
<td>9.3</td>
<td>8.0</td>
<td>5.8</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.89</td>
<td>0.48</td>
<td>0.47</td>
<td>0.76</td>
</tr>
<tr>
<td>Slope</td>
<td>0.62</td>
<td>0.65</td>
<td>0.68</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Forecasted meteorological parameters are used as inputs.

Training data: June 1\textsuperscript{st}-8\textsuperscript{th}, 2003.

Test data: June 9\textsuperscript{th}-30\textsuperscript{th}, 2003.

Forecasts from Environment Canada made at 6:00AM.

6 hour increments up to 48 hours.
6:00AM
Inputs: Present $O_3$, 12:00PM forecasts: Temp, WSP, RH

MLFE Rule Base
Predict: Ozone

6hour increments
Inputs: Predicted $O_3$, forecasts: Temp, WSP, RH
Inputs: Observed Past 6 hour $O_3$, forecasts: Temp, WSP, RH

MLFE Rule Base
Predict: Ozone

Forecasts made for 6 - 24 hour time lapse.
EDMONTON PM$_{2.5}$ Model

- Test data: January-February, 2005.
- Possible input variables:
  - Past PM$_{2.5}$ (from how far back?)
  - NO, NO$_2$, SO$_2$, O$_3$
  - Temp, RH, WSP, WDR
  - CO, PM$_{10}$, NH$_3$
EDMONTON PM$_{2.5}$ Model – Two Regimes:
Low PM$_{2.5}$ & High PM$_{2.5}$

- Past PM$_{2.5}$ (12 hours past)
- O$_3$
- RH

- Past PM$_{2.5}$ (12 hours past)
- NO$_2$
- Temp

Low PM$_{2.5}$
< 11 ug/m$^3$

High PM$_{2.5}$
> 11 ug/m$^3$
Fuzzy logic has the potential to be a very useful tool in air pollution modeling.

The MLFE model is good at tracking changes and predicting ozone concentrations.

Care must be taken in choosing the training data and tuning parameters.

The MLFE model is simple and does not require extensive computing power.

MLFE and PM$_{2.5}$ (?)
FUZZY LOGIC MODELING OF SURFACE $O_3$ AND $PM_{2.5}$

Contact:
Rachel.Mintz@ec.gc.ca

QUESTIONS?
ACKNOWLEDGEMENTS

- Dr. Svrcek and Dr. Young
- Fellow Graduate Students
- The University of Calgary and the Department of Chemical and Petroleum Engineering
- NSERC, AIF and CEERE