AQMN Optimization Using GIS Interpolation Techniques

Mujtaba Shareef, PhD, PEng, PMP
Air Quality Scientist, Alberta Energy Regulators

Air & Waste Management Association (A&WMA)
Canadian Prairies and Northern Section (CPANS)

2018 Annual Conference
May 2018

OUTLINE

- Introduction
- Network Design
- Proposed Methodology
- Case Study
- Conclusions

INTRODUCTION

- Air Pollution
- Continuous monitoring
- Optimal Network

http://healthydebate.ca/2015/11/topic/air-pollution-health-alberta-ontario

Source: http://www.oilsands.alberta.ca/air.html

LITERATURE

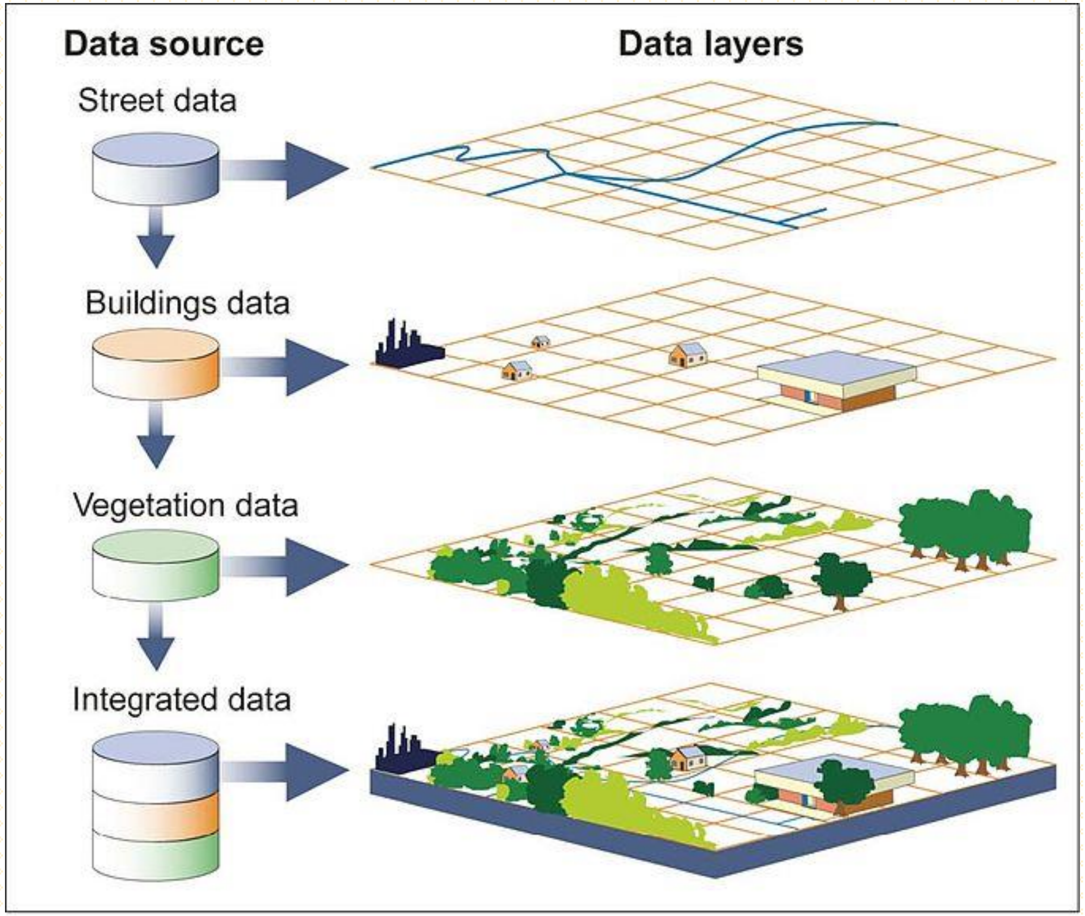
- Maximum sensitivity of collected data
- Max coverage factors
- Fishers Information Measure
  - A of measuring the amount of information that an observable random variable X carries about an unknown parameter θ of a distribution that models
- Spatial Correlation Analysis
- Expected Ambient Pollutant Dosage
  - The Dosage Monitoring Survey Design (DMSS) analyzes the dosage impact at grid receptors by dispersion modeling. High-dosage grids become potential monitoring sites.

LITERATURE

- Shannon Information
- Sphere of Influence and Figure of Merit
- Linear Programming Approach
  - Multi-attribute Utility Function
- Holistic Approach
  - Fuzzy Analytic Hierarchy Process
- Genetic Algorithm
- Principal Component Approach and Cluster Analysis
- Self Organizing Map

GEOGRAPHICAL INFORMATION SYSTEMS

- Very Popular
- Interpolation Techniques
- Automated and Data Integrated



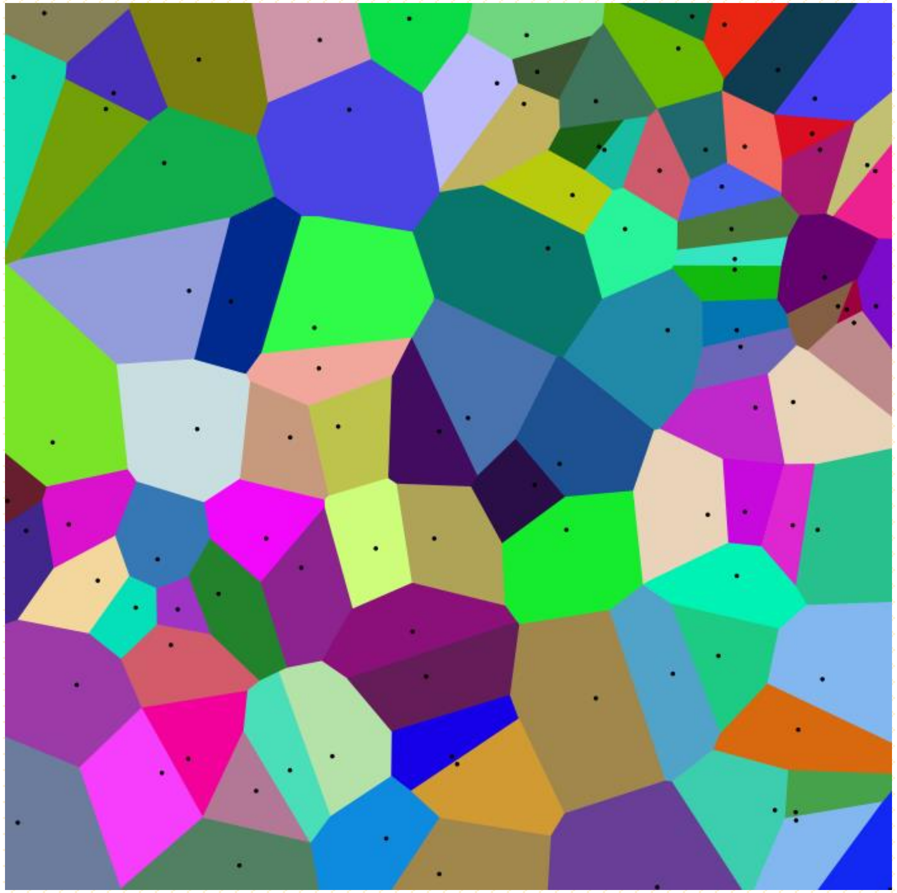
Source: GAO.

PROPOSED METHODOLOGY

- Interpolation Methods
- Performance Measures
- Station Elimination Process

INTERPOLATION METHODS

- Predicts values for cells in a rater
- Five methods
  - Inverse Distance Weighted (IDW)
    - Average from the surrounding cells
  - Spline (SPL)
    - minimizes overall surface curvature thus creating smooth surface
  - Ordinary Kriging (OK)
    - mean value is assumed constant and determined during interpolation
  - Universal Kriging (UK)
    - Data follows a trend
  - Natural Neighbor (NN)
    - finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas to interpolate a value



PERFORMANCE MEASURES

- Statistical Error – Amount by which an observation differs from its expected value
- Techniques
  - Root Mean Square Error (RMSE)
    - Frequently used
    - SD of differences between predicted and observed
    - Lower values - good prediction
  - Mean Absolute Percentage Error (MAPE)
  - Nash-Sutcliffe Equation (NSE)
    - The coefficient of efficiency
    - Indicated normalized fit of the model
    - Value ranges from $-\infty$ to 1
  - Accuracy Factor (ACFT)
    - Simple multiplicative factor
    - 1 to ∞
  - Pearson Correlation Coefficient (r)
- RMSE < 8
- MAPE < 25
- $r^2 > 0.5$

PERFORMANCE MEASURES - TECHNIQUES

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Intr_i - Obs_i\right)^2}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{Intr_i - Obs_i}{Intr_i}\right|\right)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}\left(Intr_i - Obs_i\right)^2}{\sum_{i=1}^{n}\left(Intr_i - Obs_i^{*}\right)^2}$$

$$\mathrm{ACFT} = 10^{\left(e^{\frac{\left|\log\left(\frac{Intr_i}{Obs_i}\right)\right|}{n}}\right)}$$

where $Intr_i$ = Interpolated value; Obs_i = Observed value; n = number of observations.

STATION ELIMINATION PROCESS

- Objective
  - To eliminate as many stations as possible and fill in the missing information through interpolated values
- Process
  - Eight Steps

STATION ELIMINATION PROCESS

- Step 1: Select Stations
  - Set of stations (P1, P2...) are selected to be eliminated from the vector dataset used for creating the raster.
  - A particular station or set of stations can be chosen or set of all possible stations tested in a loop.
- Step 2: Store Observed Values
  - The observed concentrations at the selected station [P ()] or stations [P1 (), P2 () …] are stored as arrays.
  - These values will later be compared with the interpolated values

STATION ELIMINATION PROCESS

- Step 3: Create Vector Layer
  - Point Layer with ‘z’.
  - Several columns depending on number of pollutants
- Step 4: Create Raster
  - Array of equally sized cells (rows and columns)
  - Five Rasters – IDW, Spline, OK, UK, NN

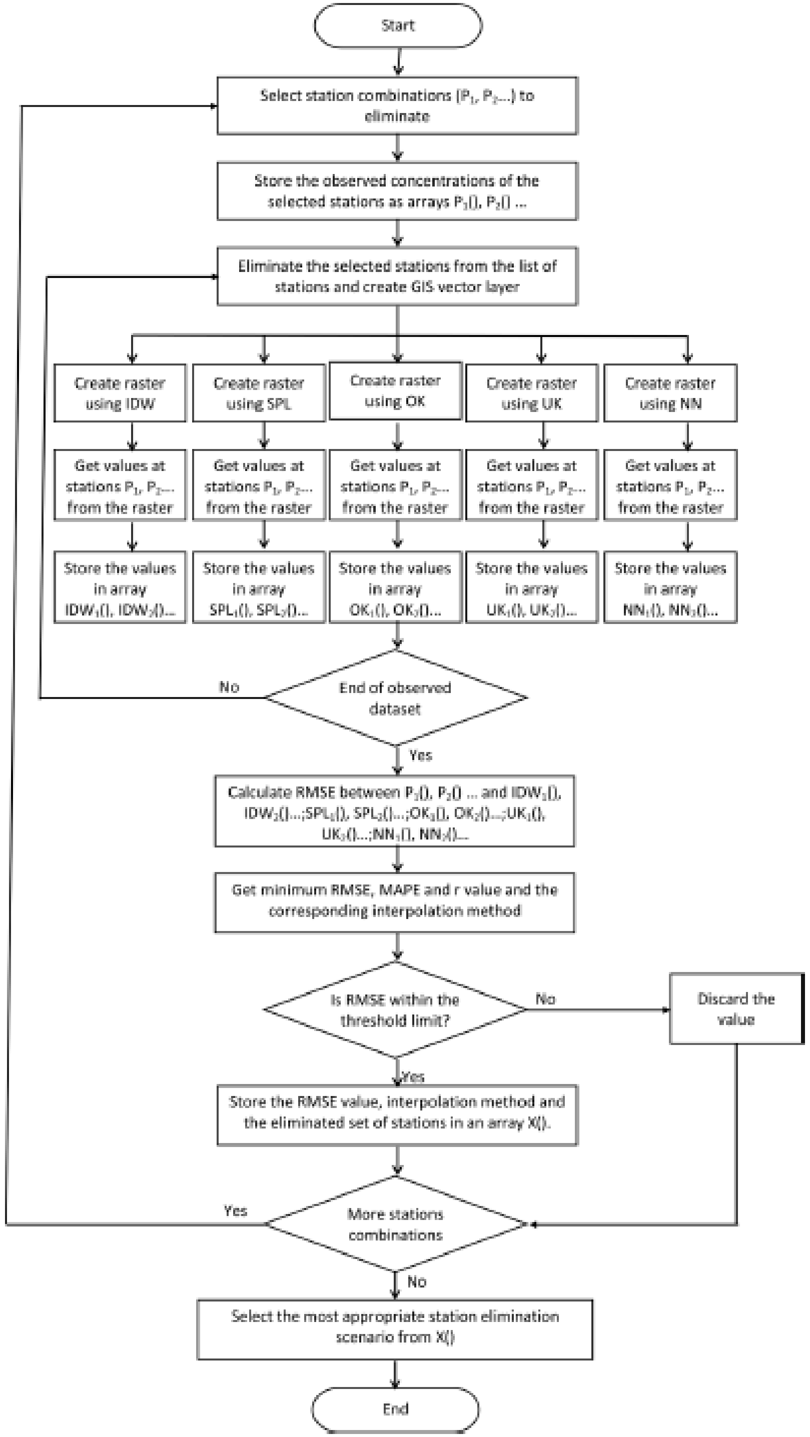
STATION ELIMINATION PROCESS

- Step 5: Extract Interpolated Value
  - Calculates the predicted concentrations
  - Values extracted from the created rasters
  - $IDW_1()$, $IDW_2()$………..$IDW_n()$
  - $SPL_1()$, $SPL_2()$…………..$SPL_n()$
  - $OK_1()$, $OK_2()$…………….$OK_n()$
  - $UK_1()$, $UK_2()$…………….$UK_n()$
  - $NN_1()$, $NN_2()$…………….$NN_n()$

STATION ELIMINATION PROCESS

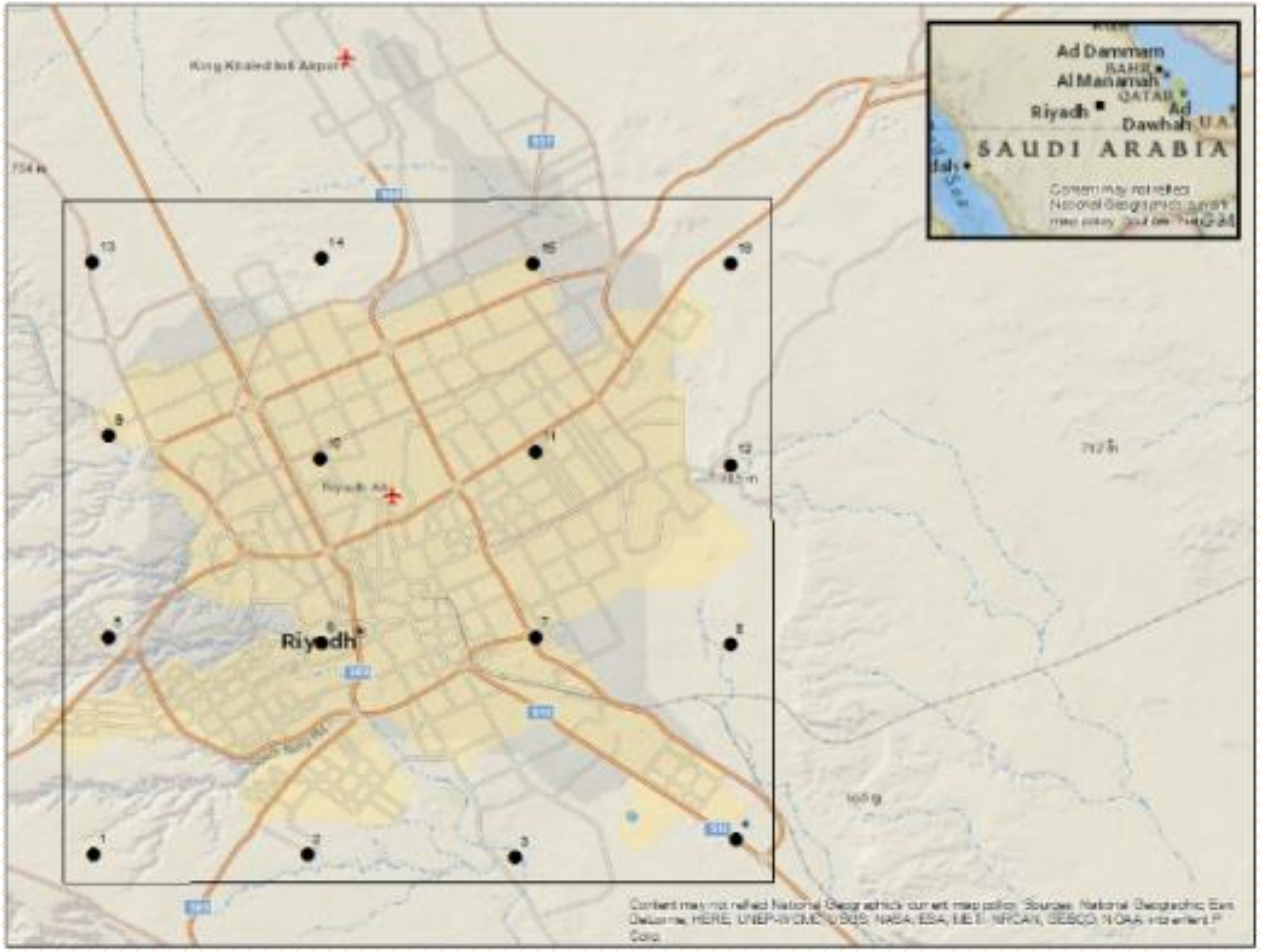
- Step 6: Measure Performance
  - Generates array of interpolated values
  - Observed vs interpolated values
  - RMSE, Bias, and Correlation Coefficient
  - Below minimum performance discarded
  - Stored in possible combinations array C()
- Step 7: Repeat Next Station
- Step 8: Find the best possible combination
  - Best from C()
  - Decision maker can choose from the list

STATION ELIMINATION PROCESS



CASE STUDY

- Study Location
  - Riyadh, Saudi Arabia
  - Sixteen Cells 12 km x km
  - Sep 2011 – Sep 2012
- Field Measurements
  - SO_2, NO_x, O_3, and CO
- Rasters
  - ESRI ArcGIS
  - Scripts



RESULTS AND DISCUSSIONS

Alberta Energy Regulator

- Assumptions - Minimum Eight(8) stations needed to produce reliable interpolated maps
- Eight (8) possible elimination scenarios

No. of stations to be eliminated	Possible combinations	No. of simulations	Approximate time of simulation (hours)
1	16	1920	0.3
2	120	14,400	2.25
3	560	67,200	10.5
4	1820	218,400	34
5	4368	524,160	81
6	8008	960,960	150
7	11,440	1,372,800	214
8	12,870	1,544,400	241

RESULTS AND DISCUSSIONS

Alberta Energy Regulator

O_3

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r^2	MAPE	NSE	ACFT
1	11	IDW	3.224	0.953	4.041	0.944	0.960
2	11, 15	IDW	3.811	0.925	5.153	0.919	0.973
3	4, 11, 15	UK	3.964	0.913	7.991	0.901	1.003
4	4, 11, 12, 15	UK	4.588	0.911	15.699	0.884	1.099
5	4, 5, 11, 15, 16	IDW	5.495	0.862	22.805	0.844	1.095
6	4, 8, 11, 12, 13, 15	UK	6.113	0.831	25.322	0.821	1.015
7	4, 5, 8, 11, 12, 15, 16	IDW	6.488	0.832	36.469	0.809	1.081
8	4, 5, 6, 8, 11, 12, 15, 16	IDW	7.018	0.817	49.610	0.775	1.138

RESULTS AND DISCUSSIONS

Alberta Energy Regulator

NO_x

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r^2	MAPE	NSE	ACFT
1	11	IDW	5.999	0.865	3.792	0.848	1.020
2	3, 11	IDW	7.470	0.808	4.006	0.731	0.961
3	3, 5, 11	UK	9.084	0.619	6.591	0.599	0.984
4	3, 5, 7, 11	UK	10.766	0.439	12.184	0.430	1.025
5	3, 5, 7, 11, 14	IDW	11.521	0.461	15.621	0.459	1.022
6	3 , 5, 7, 11, 14, 16	UK	12.119	0.438	19.839	0.438	1.019
7	3 , 5, 6, 7, 11, 14, 16	IDW	14.834	0.395	23.148	0.341	0.966
8	1, 3 , 5, 6, 7, 11, 14, 16	IDW	14.684	0.312	27.468	0.287	0.977

RESULTS AND DISCUSSIONS

Alberta Energy Regulator

SO_2

No. of station(s) to be eliminated	Optimal station(s) combination that can be eliminated	Interpol. method	RMSE	r^2	MAPE	NSE	ACFT
1	1	UK	2.155	0.323	2.155	−1.507	0.983
2	1, 4	IDW	3.991	0.446	5.925	0.421	1.029
3	1, 4, 13	UK	4.599	0.375	10.792	0.355	1.042
4	1, 4, 6, 13	IDW	4.737	0.268	15.562	0.167	1.063
5	1, 4, 6, 13, 15	IDW	5.124	0.582	25.187	0.428	1.140
6	1, 4, 6, 11, 13, 15	IDW	6.258	0.642	43.558	0.284	1.248
7	1, 4, 6, 7, 11, 13, 15	UK	7.579	0.469	60.337	0.038	1.278
8	1, 4, 6, 7, 9, 11, 13, 15	UK	9.962	0.301	69.931	0.100	1.279

RESULTS AND DISCUSSIONS

Priority MAPE

Pollutant	Stations needed to achieve MAPE < 25
O_3	1, 2, 3, 5, 6, 7, 9, 10, 14, 16
NO_x	1, 2, 4, 8, 9, 10, 12, 13, 15
SO_2	2, 3, 5, 7, 8, 9, 10, 11, 12, 14, 16
CO	1, 3, 4, 5, 6, 7, 8, 11, 12, 13, 15

RESULTS AND DISCUSSIONS

Parameters with priority as NO_x

Pollutant	RMSE	r^2	MAPE	NSE	ACFT
NO_x	14.834	0.396	23.148	0.341	0.966
CO	0.565	0.285	73.538	0.275	1.183
O_3	10.736	0.815	85.331	0.432	1.438
SO_2	37.826	0.109	198.364	−0.194	0.900

RESULTS AND DISCUSSIONS

Parameters with priority as SO_2

Pollutant	RMSE	r^2	MAPE	NSE	ACFT
SO_2	5.125	0.582	25.128	0.431	1.139
O_3	11.834	0.550	28.723	0.496	0.929
NO_x	26.512	0.063	51.687	0.045	1.101
CO	0.740	0.182	80.081	0.141	1.038

CONCLUSIONS

- Simple approach
- Existing stations are systematically eliminated
- GIS rasters
- RMSE, MAPE, r2
- Automated (scripted)
- Find optimal stations
- Need continuous data
- More parameters can be included

- Thank You