D$_{o}$MiNO - Spatial data mining exploring co-location of adverse birth outcomes and environmental variables.

Dr. Alvaro R. Osornio-Vargas

Canadian Perinatal Programs Coalition

June 23rd & 24th, 2014

Montréal, Quebec

Awarded Collaborative Health Research Program, CIHR/NSERC Grant
April 2013 – March 2016

DoMiNO is:

- Interdisciplinary and Exploratory
- Uses publicly funded data bases
- DATA MINING

Background

Current research identifies *associations between ABO and various determinants of health:*

- – social factors (e.g. poverty, stress),
- – biological factors (e.g. diabetes, infection, maternal age)
- – environmental pollutants (e.g. metals, PM_{10}, SO_2)

Growing evidence linking Urban Air Pollutants and Adverse Birth outcomes

Background

Complex problem:

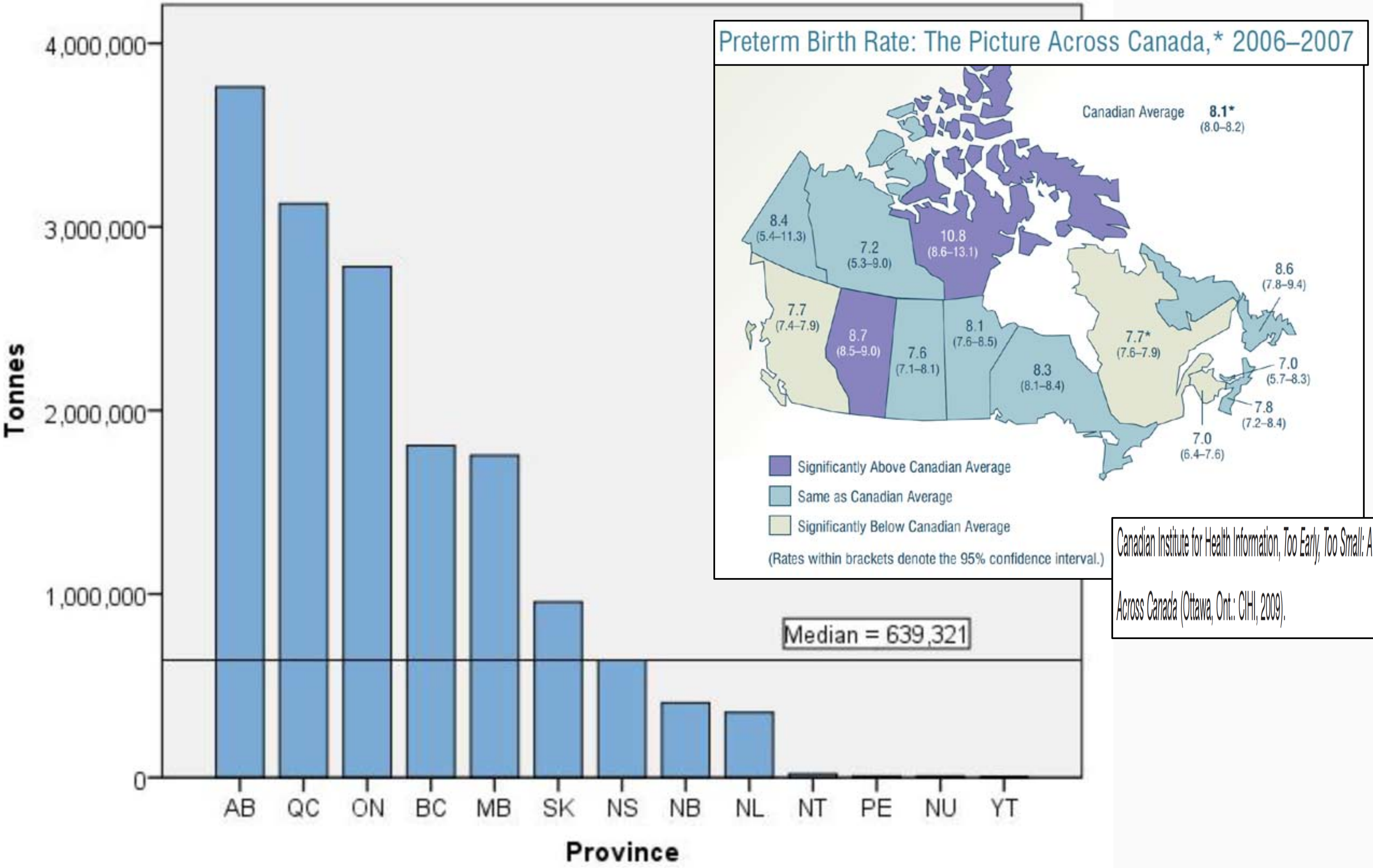
- Multiple sources of pollutants (e.g. traffic, industry).
- Interactions, dispersion, transport and fate of pollutants
- Intrinsic toxicity of pollutants
- Interactions between, social, biological, chemical and physical factors

Background

In order to advance research on links between *environmental pollutants and ABO,* methods for comprehensive assessment of the multiple variables interacting in complex ways are required.

DATA MINING

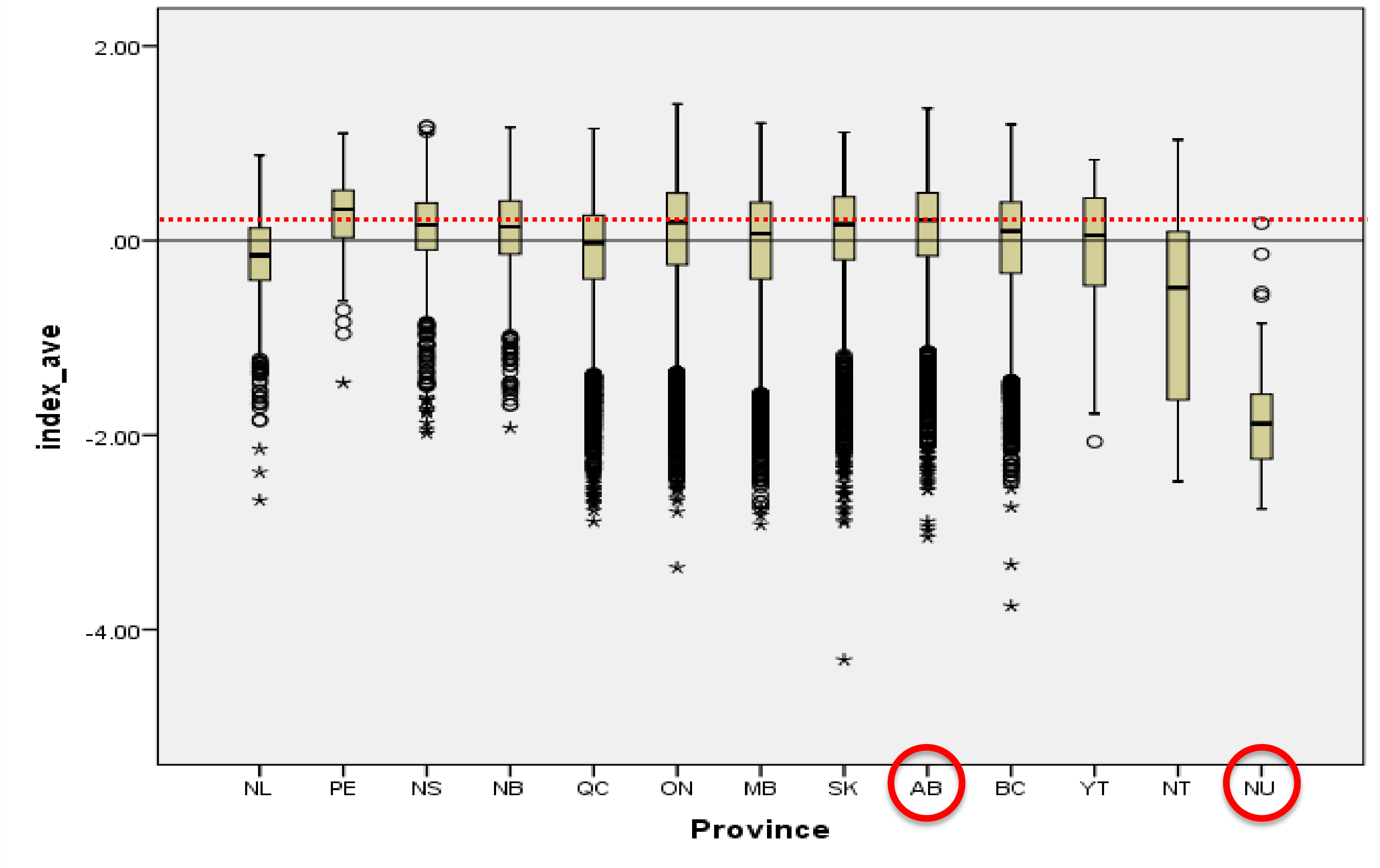
Spatial data mining exploring co-location of adverse birth outcomes and environmental variables. Osornio-Vargas, A.R.



Canadian Institute for Health Information, *Too Early, Too Small: A Profile of Small Babies Across Canada* (Ottawa, Ont.: CIHI, 2009).

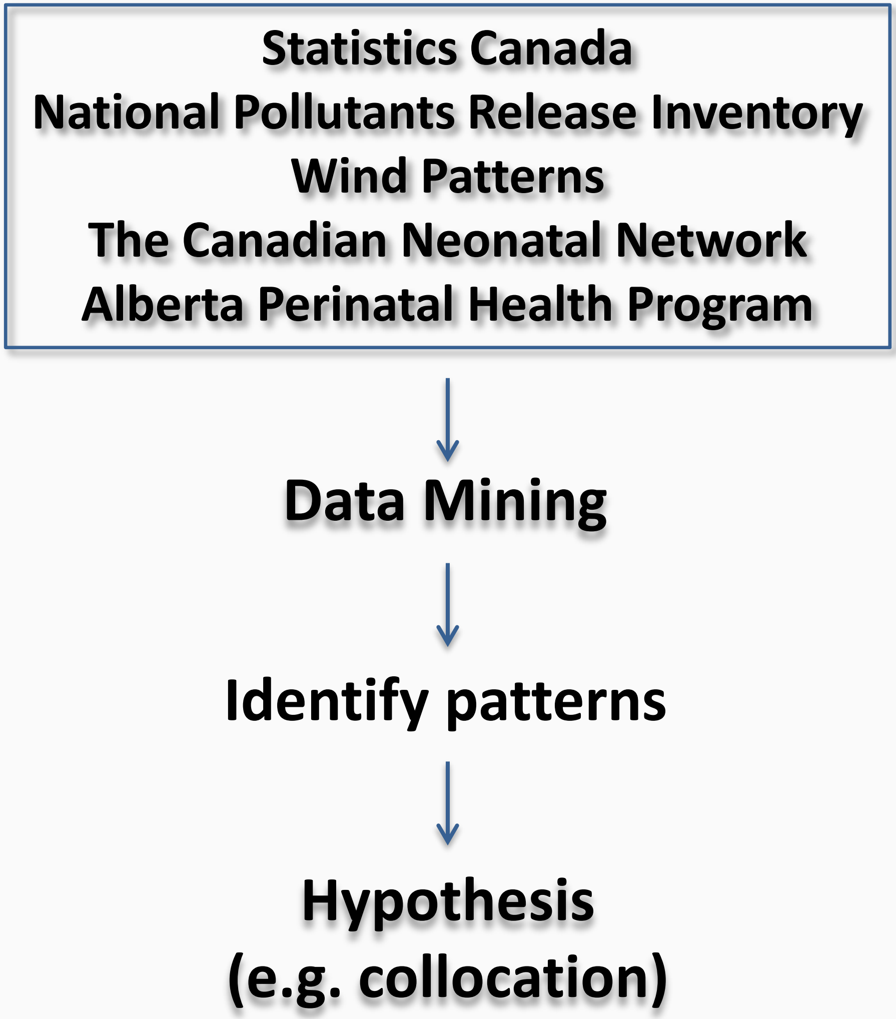
Figure 2. *Tonnes of developmental toxicants reported to NPRI as released to air in each Province of Canada, from 2006 – 2010. Numbers of chemicals released vary by Province. The line marks median emissions.*

Distribution of SES index by province and territory



Adverse birth outcomes and the environment

2002 - 2010



Statistics Canada
National Pollutants Release Inventory
Wind Patterns
The Canadian Neonatal Network
Alberta Perinatal Health Program

Data Mining

Identify patterns

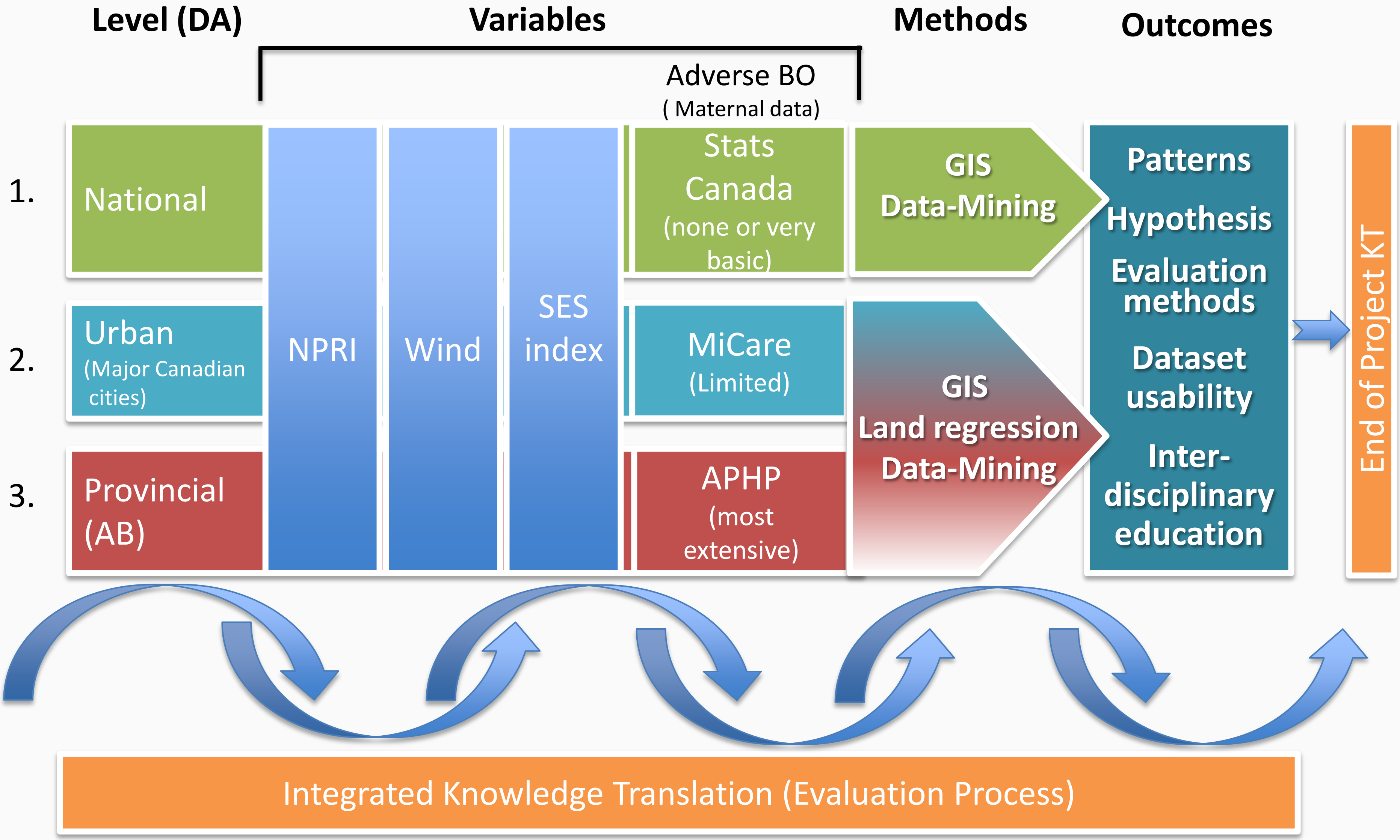
Hypothesis
(e.g. collocation)

Objective

Aims to:

- Identify patterns
- Generate hypothesis
- Spatial collocation
  - Adverse birth outcomes
    (pre-term, low birth weight and perinatal mortality)
  - Environmental Factors
    (pollution, SES)

Research Framework



Level (DA)

Variables

Methods

Outcomes

Adverse BO

(Maternal data)

1. National

2. Urban (Major Canadian cities)

3. Provincial (AB)

NPRI

Wind

SES index

Stats Canada (none or very basic)

MiCare (Limited)

APHP (most extensive)

GIS Data-Mining

GIS Land regression Data-Mining

Patterns

Hypothesis

Evaluation methods

Dataset usability

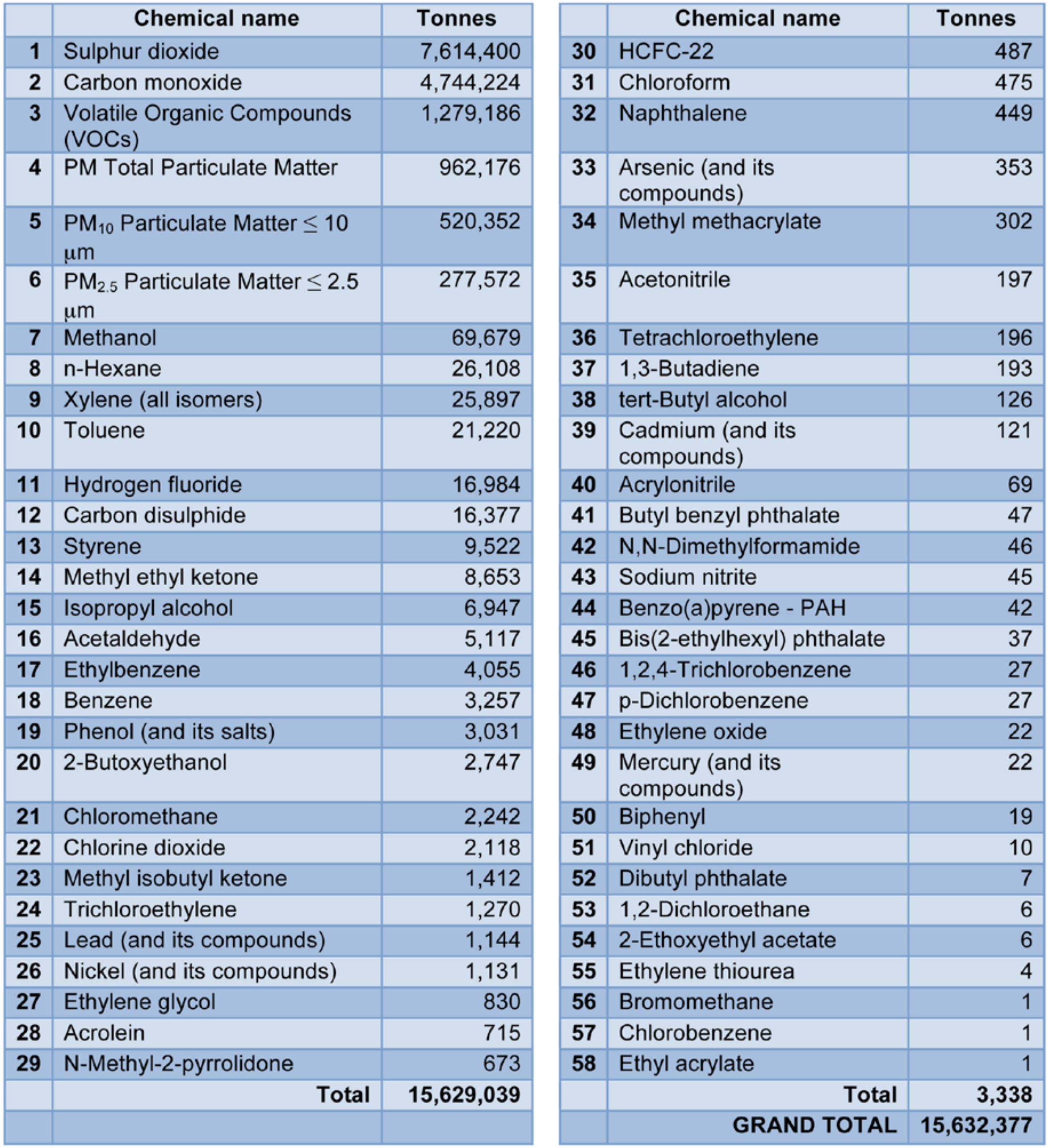
Inter-disciplinary education

End of Project KT

Integrated Knowledge Translation (Evaluation Process)

DoMiNO

Data Mining newborn outcomes



	Chemical name	Tonnes
1	Sulphur dioxide	7,614,400
2	Carbon monoxide	4,744,224
3	Volatile Organic Compounds (VOCs)	1,279,186
4	PM Total Particulate Matter	962,176
5	PM_{10} Particulate Matter ≤ 10 µm	520,352
6	$PM_{2.5}$ Particulate Matter ≤ 2.5 µm	277,572
7	Methanol	69,679
8	n-Hexane	26,108
9	Xylene (all isomers)	25,897
10	Toluene	21,220
11	Hydrogen fluoride	16,984
12	Carbon disulphide	16,377
13	Styrene	9,522
14	Methyl ethyl ketone	8,653
15	Isopropyl alcohol	6,947
16	Acetaldehyde	5,117
17	Ethylbenzene	4,055
18	Benzene	3,257
19	Phenol (and its salts)	3,031
20	2-Butoxyethanol	2,747
21	Chloromethane	2,242
22	Chlorine dioxide	2,118
23	Methyl isobutyl ketone	1,412
24	Trichloroethylene	1,270
25	Lead (and its compounds)	1,144
26	Nickel (and its compounds)	1,131
27	Ethylene glycol	830
28	Acrolein	715
29	N-Methyl-2-pyrrolidone	673
	Total	**15,629,039**
30	HCFC-22	487
31	Chloroform	475
32	Naphthalene	449
33	Arsenic (and its compounds)	353
34	Methyl methacrylate	302
35	Acetonitrile	197
36	Tetrachloroethylene	196
37	1,3-Butadiene	193
38	tert-Butyl alcohol	126
39	Cadmium (and its compounds)	121
40	Acrylonitrile	69
41	Butyl benzyl phthalate	47
42	N,N-Dimethylformamide	46
43	Sodium nitrite	45
44	Benzo(a)pyrene - PAH	42
45	Bis(2-ethylhexyl) phthalate	37
46	1,2,4-Trichlorobenzene	27
47	p-Dichlorobenzene	27
48	Ethylene oxide	22
49	Mercury (and its compounds)	22
50	Biphenyl	19
51	Vinyl chloride	10
52	Dibutyl phthalate	7
53	1,2-Dichloroethane	6
54	2-Ethoxyethyl acetate	6
55	Ethylene thiourea	4
56	Bromomethane	1
57	Chlorobenzene	1
58	Ethyl acrylate	1
	Total	**3,338**
	GRAND TOTAL	**15,632,377**

Table I :Total amounts of developmental toxicants reported to NPRI as released to air in Canada in 2006 – 2010

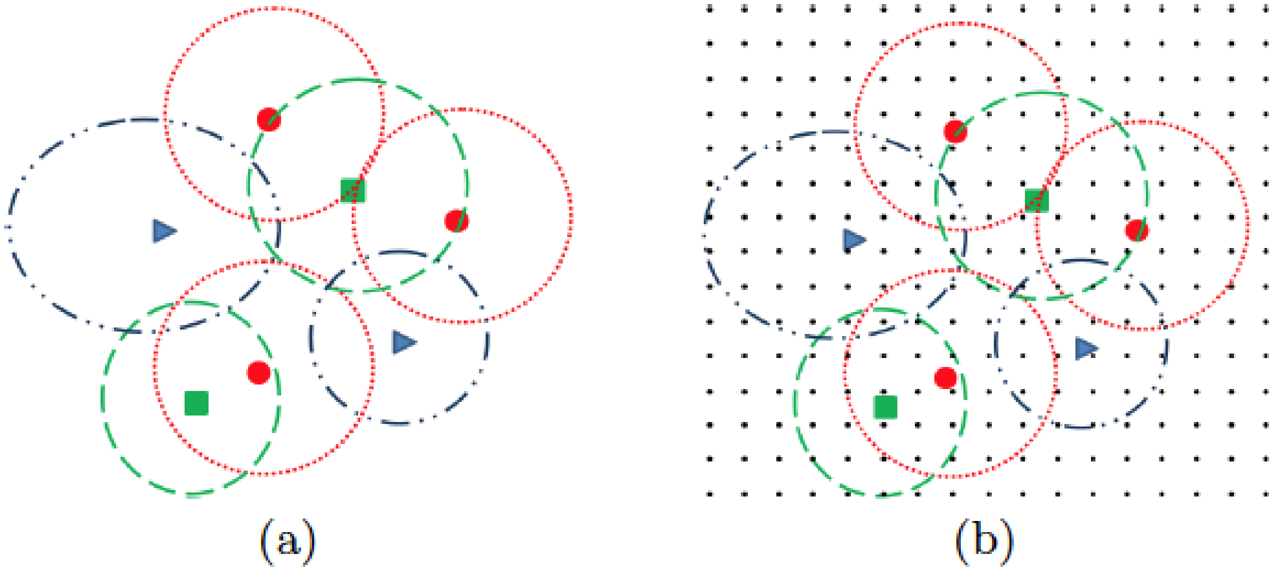
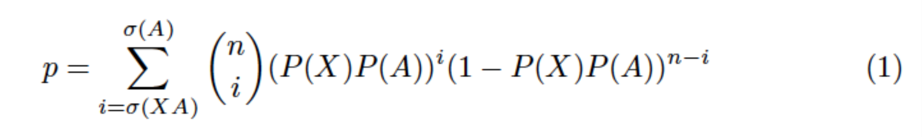
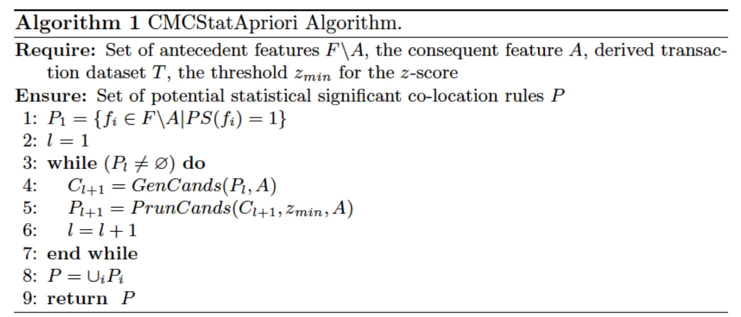


Fig. 1. Transactionization step: (a) An example spatial dataset with point feature instances and their buffers; (b) Grids imposed over the space.



$$p = \sum_{i=\sigma(XA)}^{\sigma(A)} \binom{n}{i} (P(X)P(A))^i (1 - P(X)P(A))^{n-i} \quad (1)$$



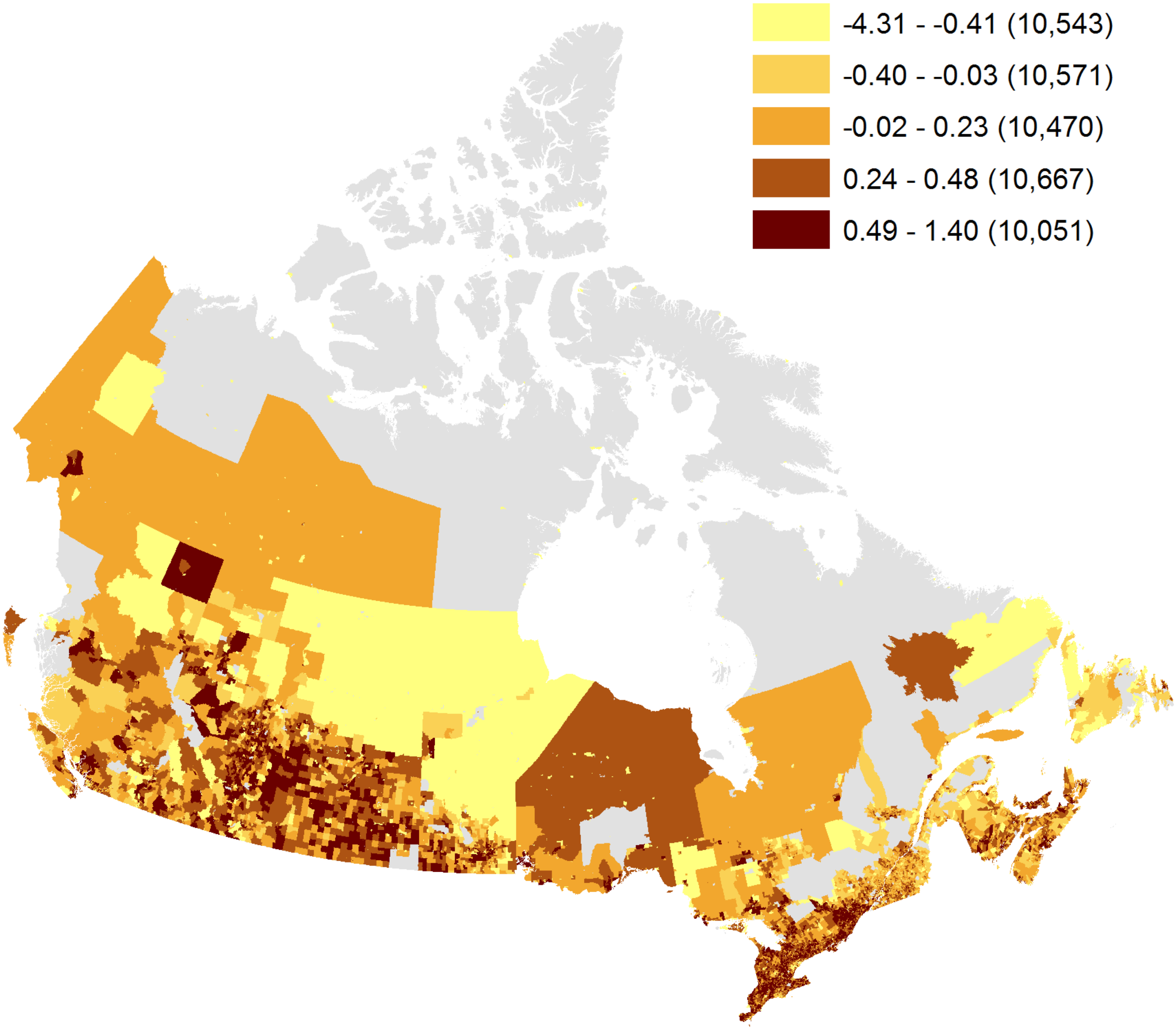
Algorithm 1 CMCStatApriori Algorithm.

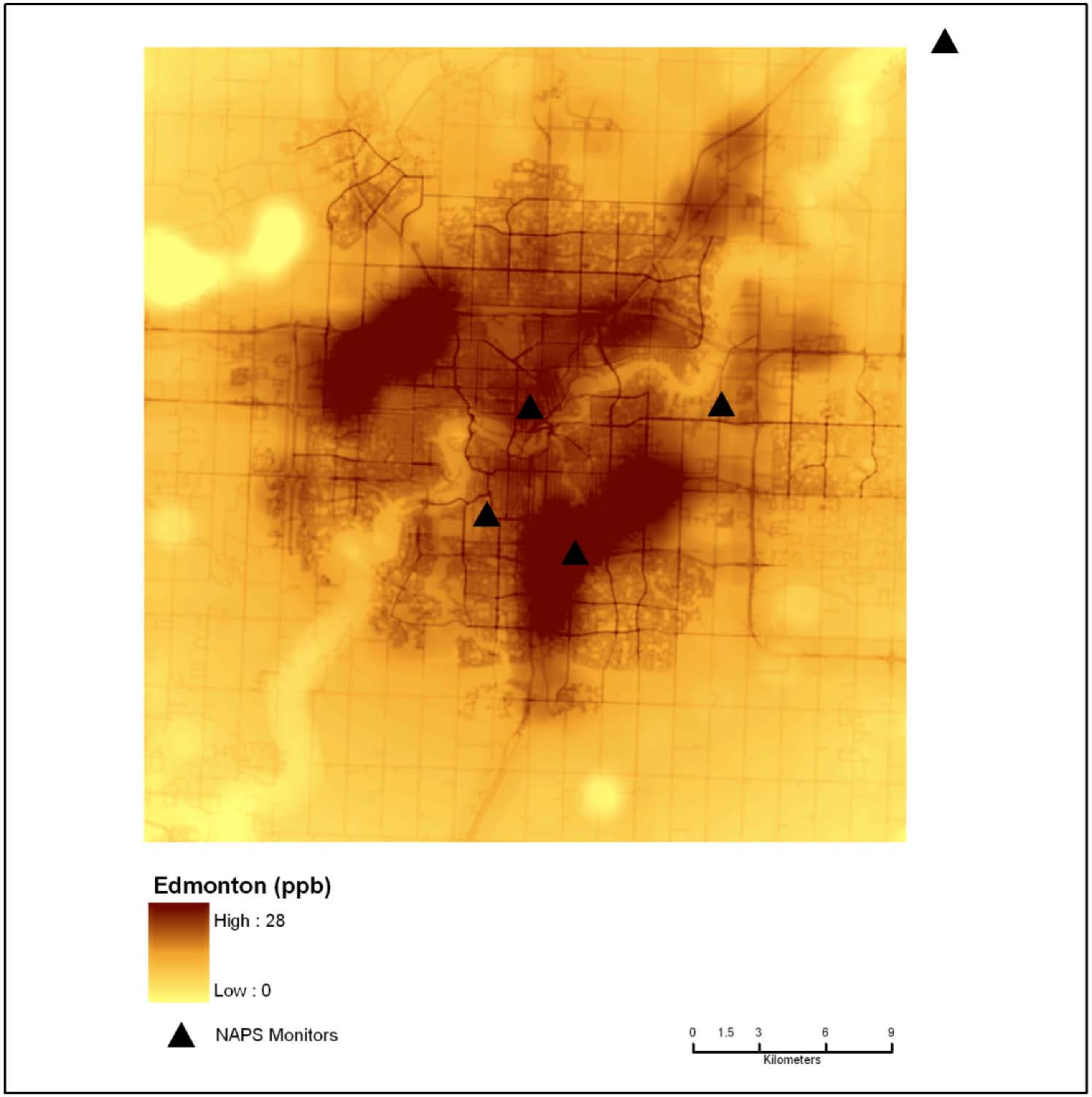
Require: Set of antecedent features $F \backslash A$, the consequent feature A, derived transaction dataset T, the threshold z_{min} for the z-score

Ensure: Set of potential statistical significant co-location rules P

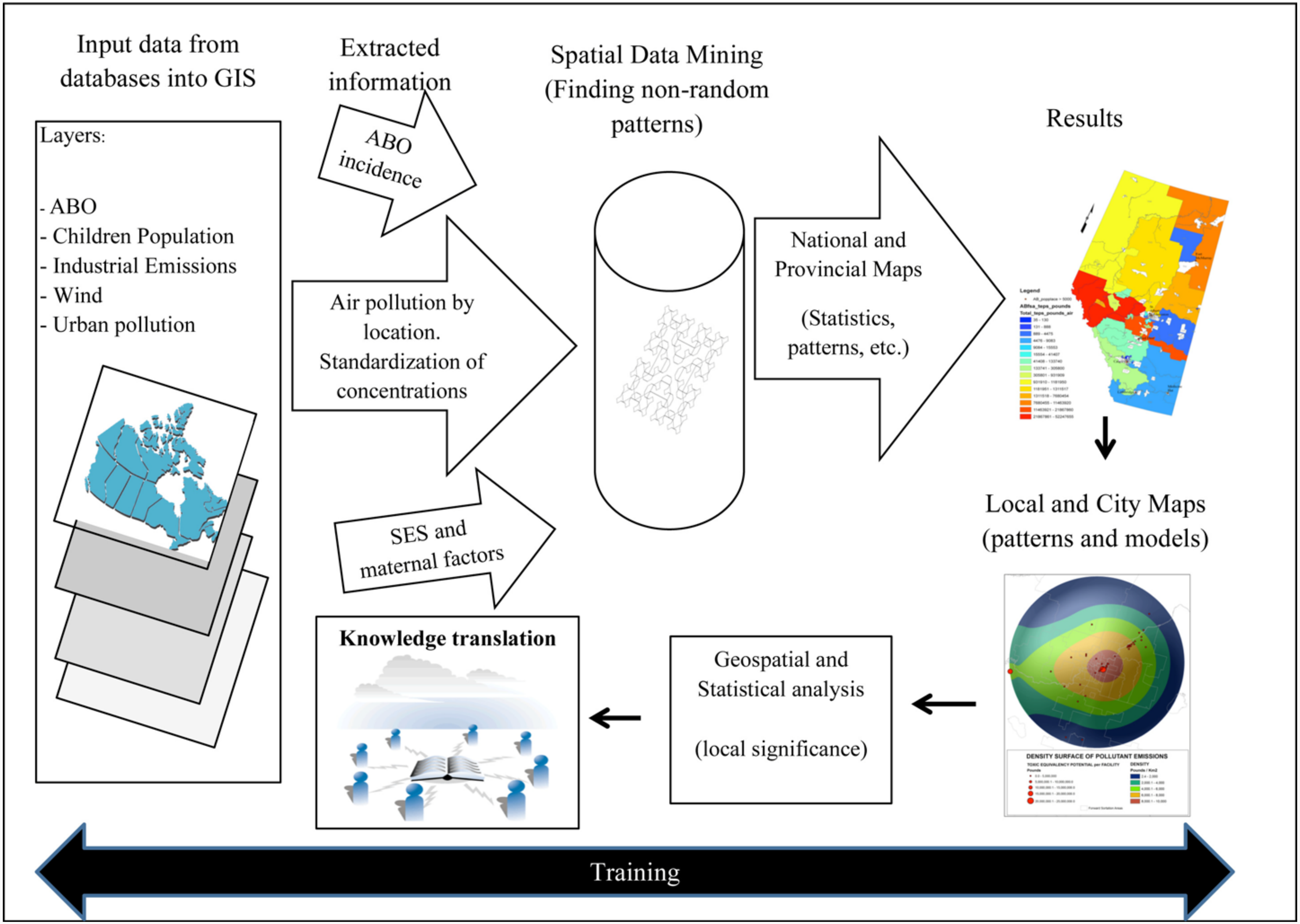
```
$P_1 = \{f_i \in F \backslash A | PS(f_i) = 1\}$
$l = 1$
while ($P_l \neq \varnothing$) do
    $C_{l+1} = GenCands(P_l, A)$
    $P_{l+1} = PrunCands(C_{l+1}, z_{min}, A)$
    $l = l + 1$
end while
$P = \cup_i P_i$
return $P$
```

SESIndex at the DA level





R.W. Allen et al. / Atmospheric Environment 45 (2011) 369-378



Input data from databases into GIS

Layers:

- ABO
- Children Population
- Industrial Emissions
- Wind
- Urban pollution

Extracted information

ABO incidence

Air pollution by location. Standardization of concentrations

SES and maternal factors

Spatial Data Mining (Finding non-random patterns)

National and Provincial Maps

(Statistics, patterns, etc.)

Results

Local and City Maps (patterns and models)

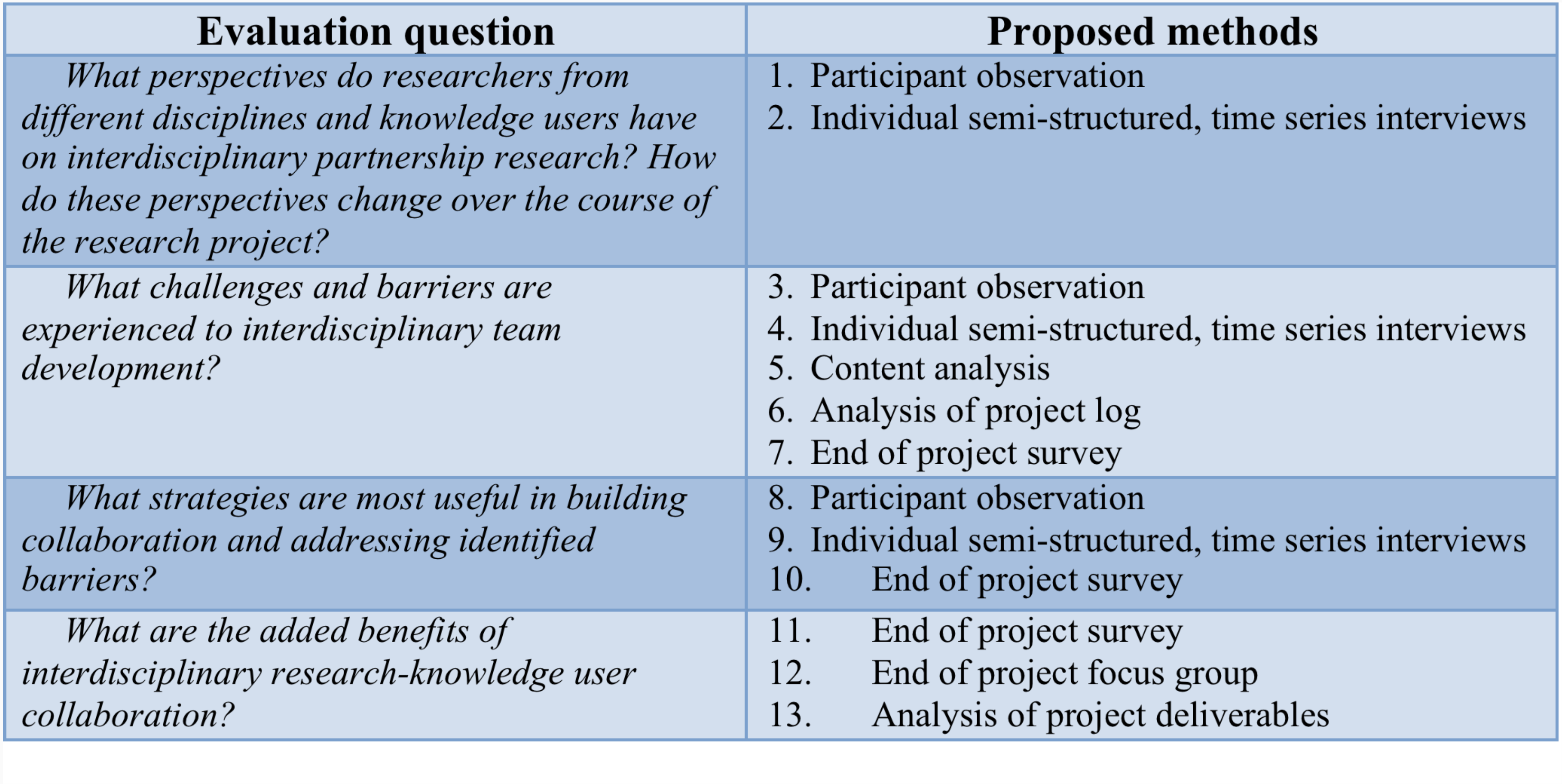
Geospatial and Statistical analysis

(local significance)

Knowledge translation

Training

Table II. Evaluation of the integrated KT plan.



Evaluation question	Proposed methods
What perspectives do researchers from different disciplines and knowledge users have on interdisciplinary partnership research? How do these perspectives change over the course of the research project?	1. Participant observation 2. Individual semi-structured, time series interviews
What challenges and barriers are experienced to interdisciplinary team development?	3. Participant observation 4. Individual semi-structured, time series interviews 5. Content analysis 6. Analysis of project log 7. End of project survey
What strategies are most useful in building collaboration and addressing identified barriers?	8. Participant observation 9. Individual semi-structured, time series interviews 10. End of project survey
What are the added benefits of interdisciplinary research-knowledge user collaboration?	11. End of project survey 12. End of project focus group 13. Analysis of project deliverables

Team

University of Alberta

Faculty of Medicine & Dentistry

Dr. Osornio-Vargas*	Principal Investigator
Dr. Irena Buka	Children’s environmental health
Dr. Khalid Aziz	Neonatology
Dr. Manoj Kumar	Neonatology
Dr. Sue Chandra	Obstetrics & Gynecology
Osnat Wine	Knowledge Translation
Emily Chan	Socioeconomic variables

Computing Sciences

Dr. Osmar Zaiane*	Principal Investigator
Jundong Li	Data mining
Dr. Dr. Sajib Barua	Data mining

School of Public Health

Dr. Sarah Bowen	Knowledge Translation
Dr. Yan Yuan	Biostatistics
Dr. Yutaka Yasui	Biostatistics
Jesus Serrano	Geostatistics

Faculty of Sciences

Charlene Nielsen	Geostatistics

Carlton University

Department of Health Sciences,

Dr. Paul Villeneuve	Epidemiology

University of Victoria

Interdisciplinary Studies

Dr. Laura Arbour	Paediatrics and Genetics
Anders Erickson	GIS

Oregon State University

School of Biological & Population Health Sciences

Dr. Perry Hystad	Spatial exposure assessment

CAREX

Dr. Eleanor Setton	Exposure Assessment
Dr. Paul Demers	Epidemiology

CIHR *Maternal-Infant Care (MiCare)* Program

Dr. Prakeshkumar Shah	Neonatology

Knowledge Users

Health Canada

Dr. David Stieb	Epidemiology
Dr. Phil Blagden	Science Advisor

Alberta Perinatal Health Program

Nancy Aelicks

Canadian Partnership for Children's Health & Environment

Erica Phipps

Thank you

osornio@ualberta.ca

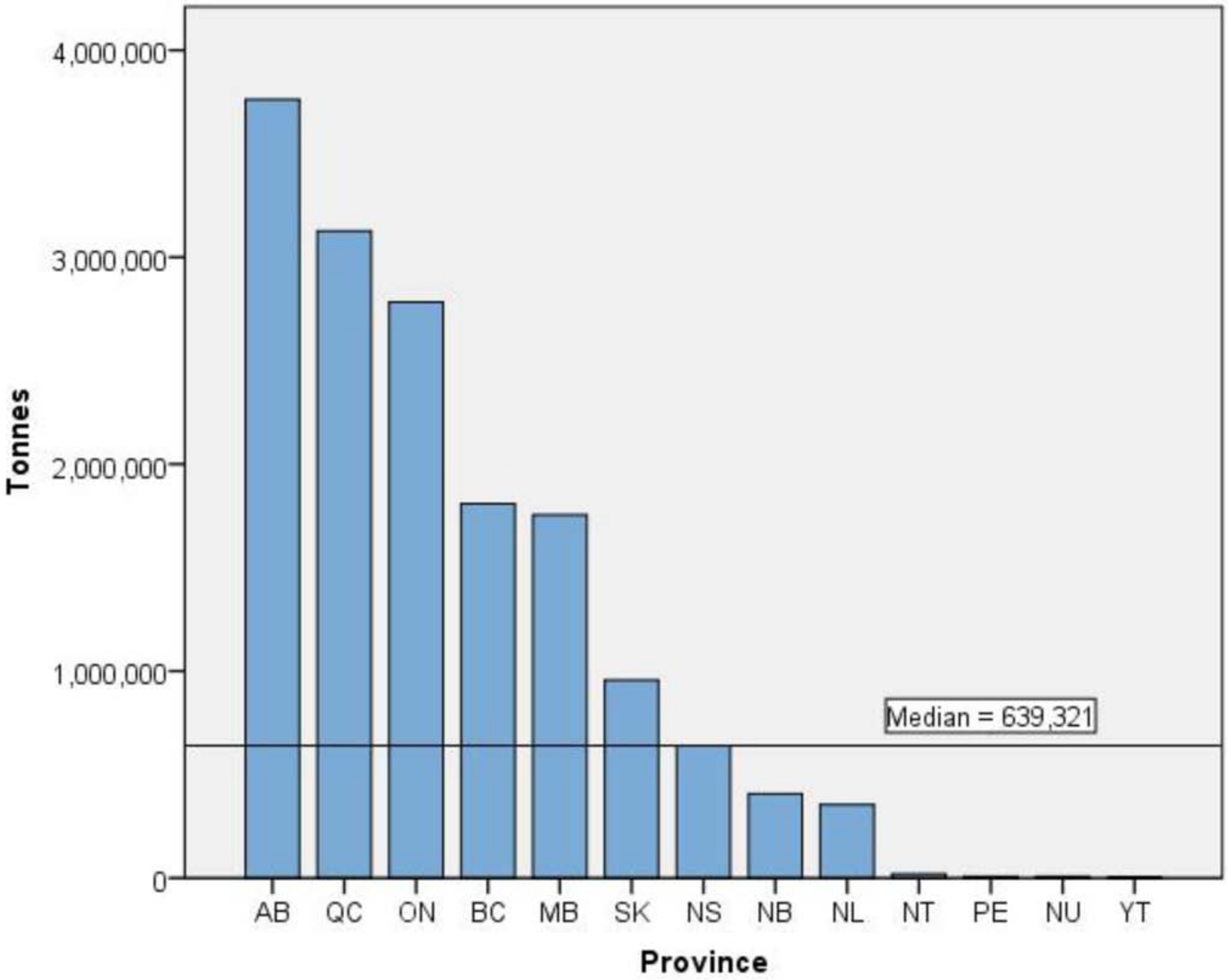
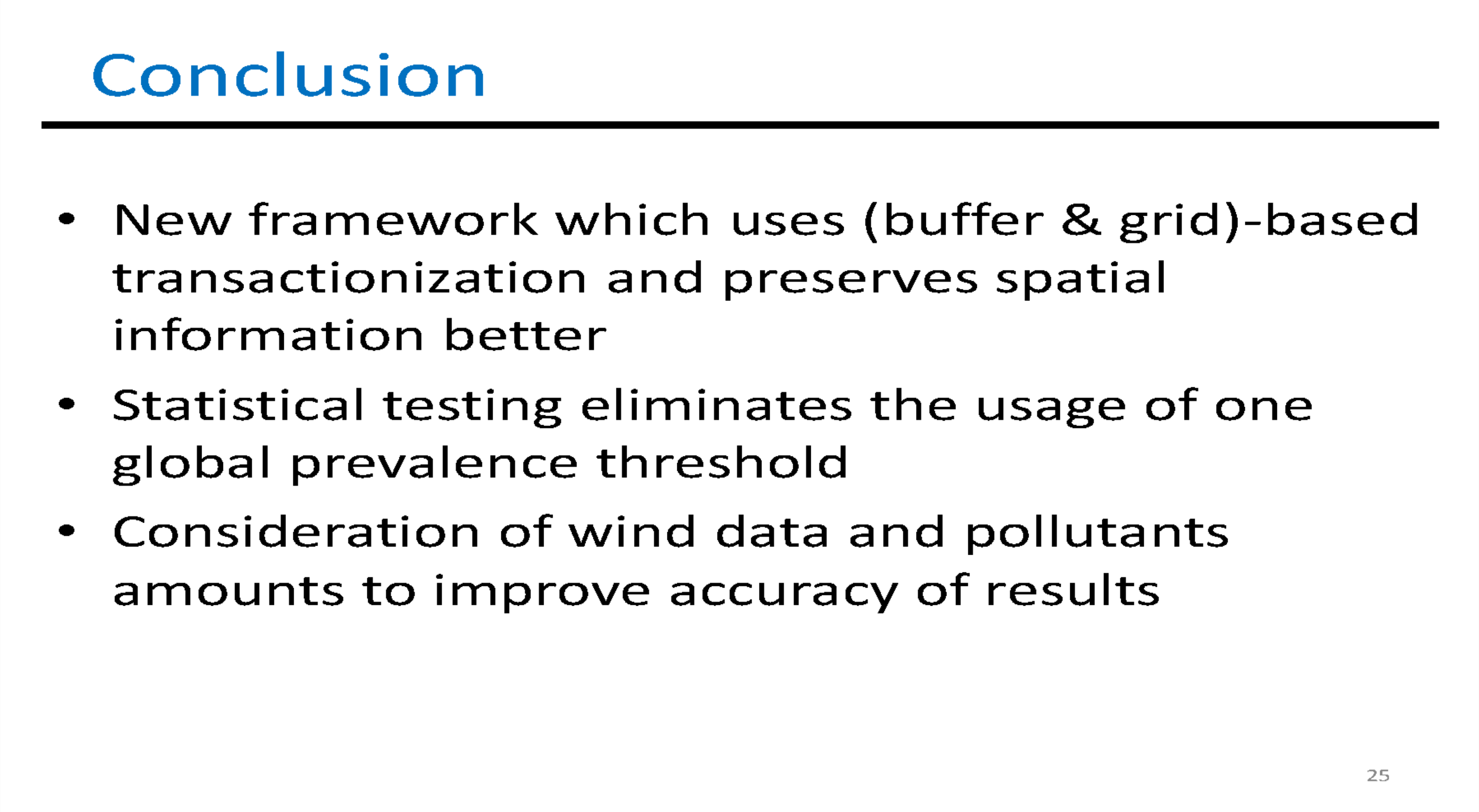


Figure 2. *Tonnes of developmental toxicants reported to NPRI as released to air in each Province of Canada, from 2006 – 2010. Numbers of chemicals released vary by Province. The line marks median emissions.*

Data Mining



Conclusion

- New framework which uses (buffer & grid)-based transactionization and preserves spatial information better
- Statistical testing eliminates the usage of one global prevalence threshold
- Consideration of wind data and pollutants amounts to improve accuracy of results