Modelling and Classifying Random Phenomena

Colin Bellinger

School of Computer Science
Carleton University
Ottawa, Canada

cbelling@scs.carleton.ca

26 January, 2010

Overview

- The Comprehensive Test Ban Treaty
- Motivation
- The atmosphere and the major dispersion processes
- Modelling atmospheric dispersion
- The simulation process
- Preliminary results

The Comprehensive Test Ban Treaty (CTBT)

The CTBT is a United Nations treaty which will **bans all nuclear explosions in the environment** when it enters into force.
http://www.ctbto.org/

CTBT - Problem Statement

Develop and deploy a verification regime capable of ensuring the integrity of the CTBT by the time it enters into force.

Choices inherent in the development of a verification strategy.

1. Feature selection
2. Procedure for the collection and/or measurement of features
3. Receptor placement
4. Classification strategy

CTBT - Design decisions

(1) Feature selection: ${}^{131m}Xe, {}^{133}Xe, {}^{133m}Xe$ and ${}^{135}Xe$
  - selected for intermediate rates of decay
  - due to their property of being inert

(2) Air sampling equipment
  - sampling occurs over 12 or 24 hours
  - filter is cooled
  - followed by analysis which produces a gamma ray spectrum
  - feature vector extraction

CTBT - Design decisions

(3) Receptor location is largely an open question
  - ideally, elevated and subject to regular wind
  - the form of the global network which maximizes the likelihood of detection is not entirely clear

(4) Classification
  - input feature vector
  - output detonation decision
  - complicating factors: radioxenon emitted from medical isotope production facilities and nuclear power generation

Motivation

- Health Canada Dataset
  - real (background) data collected from five receptor sites
  - infused with artificial explosion data
- Limitations
  - explosion instances greatly outnumber background instances
  - overall limited supply

Related Work

- Data generation is particularly useful when:
  - Limited or, no, "real" data available for one or more classes
  - Added flexibility to increase understanding of classifier and improve confidence in future performance
- Both motivating factors in terms of CTBT verification.
  - limited/no "real" explosion data
  - HC background data under-represented - particularly a concern for one-class learning
  - Developed classifiers will be deployed in a highly variable environment
    - Manipulations to the generation process will provide insight into classifier behaviour during evolving conditions. - ex. MIs come on- or go off-line, change in global wind patterns etc. -
- see [Dietterich97, Alaiz08, Langley94, Zhu04]

Atmospheric Processes

- Fate of pollutants affected by atmospheric flows
  - byproduct of energy exchange
  - interplay between lower atmosphere and the earth's surface

Modelling the Atmosphere

- divided into sections based on vertical temperature gradient
- pollutants generally emitted into the troposphere
    - lowest layer
    - rises approx. 15km
    - contains atmospheric boundary layer (ABL) and convective boundary layer (CBL)

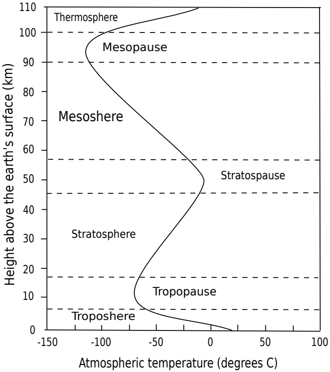


Figure: Vertical temperature profile of the atmosphere (recreated from [Cooper03].)

Atmospheric Processes

- Atmospheric boundary layer
  - lowest portion of the troposphere
  - motion suffers from frictional drag
  - resulting energy exchange affects a temperature and moisture profile
  - produces convective eddies in the above CBL
- Convective boundary layer
  - bounded below by ABL and on top by inversion layer
  - upper bound is dependent upon meteorological conditions and has a significant effect on surface pollution levels

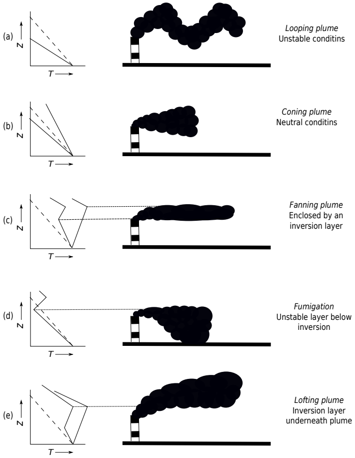
Atmospheric Processes

- Dispersion processes
  - advection: transfers pollution in the direction of mean wind
  - diffusion: shifts concentration levels down the gradient scale

  - Advection
    - limited by frictional forces resulting from surface roughness
    - wind speed increases significantly in first 10 metres
    - surface effect dissipates beyond 1 kilometre
  - Diffusion
    - random process, causes eddies to exchange and mix with neighbouring parts of atmosphere
    - largely dependent upon surface roughness and buoyancy.
    - stability term can be used to classify the degree to which mixing occurs
    - stable: little mixing
    - unstable: well mixed

Atmospheric Processes

- Effect of mixing and layering on plume [Beychok94]
    - in the inset z vs T diagram, the dashed line represents dry adiabatic lapse rate and the solid line represents measured lapse rate.
    - out is the resulting plume



Modelling the Atmosphere

Frames of reference

- Eulerian:
  - views atmosphere from a fixed point
  - inspects parcels of air as they move past
- Lagrangian:
  - describes behaviour wrt moving fluids
  - point of reference within an advecting parcel

Modelling the Atmosphere

Lagrangian particle models

- mathematically follow pollutants as they disperse by:
    - applying calculations of the statistical trajectory of parcels
    - random walk approach
- previously used to model radionuclide dispersion by [Izraehl90, Desiato92]

Modelling the Atmosphere

K-theory

- adheres to the hypothesis of gradient transfers
    - ie. assumes a shift down the gradient scale
- provides a comprehensive description of atmospheric dispersion
- example of its application can be seen in [Lauritzen99, Maccracken78]

Modelling the Atmosphere

The main equation of K-theory:

$$\frac{D\overline{\chi}}{Dt} = \frac{\partial}{\partial x}\left(K_x \frac{\partial\overline{\chi}}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y \frac{\partial\overline{\chi}}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z \frac{\partial\overline{\chi}}{\partial z}\right) \qquad (1)$$

where:

t:	time
$\overline{\chi}$:	the mean concentration of the pollutants
$K_{x,y,z}$:	eddie diffusivity in the three coordinate directions
$\frac{D\overline{\chi}}{Dt}$:	the Lagrangian time derivative

accounts for diffusion in the three component directions (anisotropic)

Modelling the Atmosphere

Assume isotropic diffusion:

- this assumes diffusion is constant and independent of spacial direction
- resulting eq'n is analogous to Fick's law of molecular diffusion

$$\frac{D\overline{\chi}}{Dt} = K\nabla^2 x \tag{2}$$

where:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

and

$$K = K_x = K_y = K_z$$

- this can be used to model diffusion in 1-, 2- or 3-dimensions

Modelling the Atmosphere

1-dimensional Fickian diffusion eq'n:

$$\frac{\partial \overline{\chi}}{\partial t} + \overline{u}\frac{\partial \overline{\chi}}{\partial x} = K\frac{\partial^2 \overline{\chi}}{\partial x^2} \tag{3}$$

where:

$$\overline{v} = \overline{w} = 0$$

and

- the first term represents the rate of change of $\overline{\chi}$ at a selected fixed point in space
- the second term represents the advection of the pollutant at velocity $\overline{u}$

Modelling the Atmosphere

Gaussian puff model:

- sol'n to Fickian diffusion equation
- models diffusion from an instaneous point source of emission strength Q
- assume mean concentration of dispersing pollutant is forms a Gaussian distribution

$$\overline{\chi}(x, y, z, t) = \frac{Q}{(4\pi t)^{\frac{3}{2}}(K_x K_y K_z)^{\frac{1}{2}}} exp\left[-\left(\frac{(x-\overline{u}t)^2}{4K_x t} + \frac{y^2}{4K_y t} + \frac{z^2}{4K_z t}\right)\right]$$

Modelling the Atmosphere

Boundary conditions Gaussian puff model:

1. $\overline{\chi} \rightarrow 0$ *as* $t \rightarrow \infty$ for all coordinates $-\infty < x, y, z < \infty$
2. $\overline{\chi} \rightarrow 0$ *as* $t \rightarrow 0$ for all coordinates except where $x, y, z = 0$
3. $\int_{-\infty}^{\infty} \overline{\chi} dxdydz = Q$

- [Oliveira98] applied a variant of the puff model to predict the dispersion of radionuclides from a nuclear power plant in Brazil

Modelling the Atmosphere

Gaussian plume model:

- models diffusion from a continuous point source (Q), emitted from an elevated industrial stack
- infinite number of puffs superimposed on each other
- mathematically speaking, integrate with respect to time
- as a matter of convenience, diffusion along x-axis is ignored

$$\chi(x, y, z, t) = \frac{Q}{2\pi\sigma_y\sigma_z\overline{u}} exp\left(-\left(\frac{y^2}{2\sigma_y^2} + \frac{z^2}{2\sigma_z^2}\right)\right)$$

Modelling the Atmosphere

Gaussian plume model:

- account for reflection at surface

$$\chi(x,y,z,t) = \frac{Q}{2\pi\sigma_y\sigma_z\overline{u}} exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left[exp\left(-\frac{(z-H)^2}{2\sigma_z^2}\right) + exp\left(-\frac{(z+H)^2}{2\sigma_z^2}\right)\right]$$

where:

h is the height of the plumes centreline

- in much the same way, reflection at an inversion layer can be accounted for
- for some further examples of the application of Gaussian models see [AlKhayat92, Simmonds93, Lyons90]

Simulating Dispersion

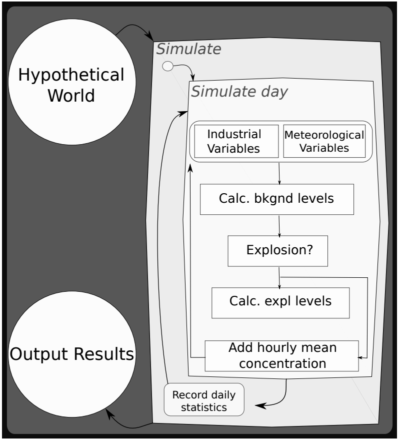
Objective:

- Generate a dataset containing a series of radioxenon measurements
- Receptor-specific datasets contain feature vectors of:
  - cumulative quantity of radioxenon measured over 12 or 24 hours
  - class label (background or explosion)

Simulating Dispersion

1. Define hypothetical world
2. Simulation for j=1:n days
    (i) For each day, simulate i=1:24 hours
        - generate Gaussian random variables about the means
        - calculate background radioxenon levels
        - if explosion, added expls levels to bkgnd levels
        - add hourly mean to cumulative daily count

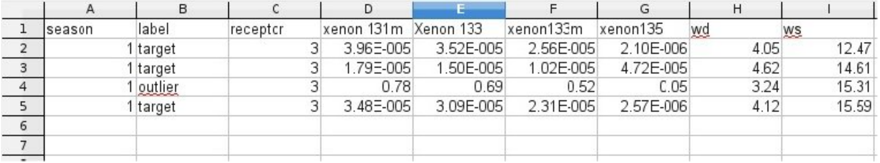
    (ii) Record daily value
3. Output dataset



Simulating Dispersion

Sample dataset:

- Important features are xenon 131m, 133, 133m and 135
- Considerations season, wind speed and direction

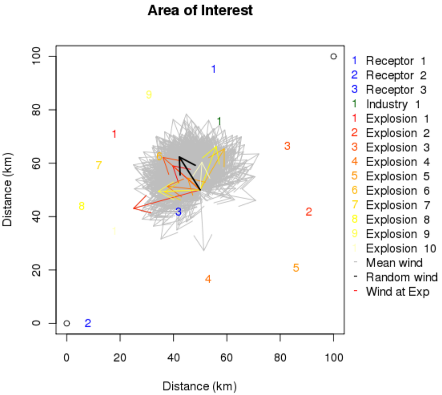


	A	B	C	D	E	F	G	H	I
1	season	label	receptor	xenon 131m	Xenon 133	xenon133m	xenon135	wd	ws
2	1	target	3	3.96E-005	3.52E-005	2.56E-005	2.10E-006	4.05	12.47
3	1	target	3	1.79E-005	1.50E-005	1.02E-005	4.72E-005	4.62	14.61
4	1	outlier	3	0.78	0.69	0.52	0.05	3.24	15.31
5	1	target	3	3.48E-005	3.09E-005	2.31E-005	2.57E-006	4.12	15.59
6									
7									

Simulating Dispersion

Sample map:

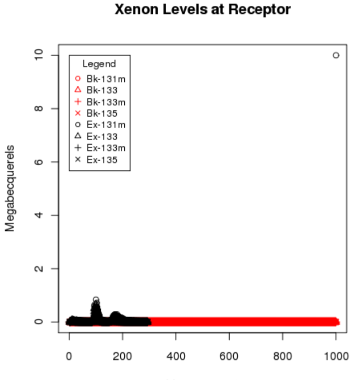
- 1 industrial emitter (green)
- 3 receptors (blue)
- 10 explosions (heat colours)



Simulating Dispersion

plotted results for receptor 2:

- background data in red
- explosions in black
- generally up-wind from industry → low background levels
- two peaks (expl 4 and expl 5)



Preliminary Results

Classifier	**Class**	**TPR**	**FPR**	**AUC**
MLP	target	0.845	0	0.896
	outlier	1	0.155	0.896
J48	target	0.997	0	0.990
	outlier	1	0.003	0.990
IBK	target	0.995	0	0.998
	outlier	1	0.005	0.998
NB	target	0.039	0.1	0.754
	outlier	0.990	0.964	0.754
CDCPE [Hempstalk08]	target	0.902	0.013	0.650
	outlier	0.087	0.098	0.650

Conclusion

- Examined strategies for modelling atmospheric dispersion
- In the spirit of the CTBT
    - applied a Gaussian assumption to model the dispersion of radioxenon
    - generate background noise and random Phenomena
- Utilized Weka to classify the preliminary dataset

TAH Al-Khayat, B. Van Eygen, CN Hewitt, and M. Kelly.
Modelling and measurement of the dispersion of radioactive emissions from a nuclear fuel fabrication plant in the U. K.
Atmospheric environment. Part A, General topics, 26(17):3079–3087, 1992.

R. Alaız-Rodrıguez and N. Japkowicz.
Assessing the Impact of Changing Environments on Classifier Performance.
In *Advances in artificial intelligence: 21st Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2008, Windsor, Canada, May 28-30, 2008: proceedings*, page 1324. Springer-Verlag New York Inc, 2008.

M.R. Beychok.
Fundamentals of stack gas dispersion.
Milton R Beychok, 1994.

J.R. Cooper, K. Randle, and R.S. Sokhi.
Radioactive Releases in the Environment, Impact and Assessment.
John Wiley & Sons Canada Ltd, 2003.

F. Desiato.
A long-range dispersion model evaluation study with chernobyl data.
Atmospheric Environment. Part A. General Topics, 26(15):2805 – 2820, 1992.

T.G. Dietterich, R.H. Lathrop, and T. Lozano-Pérez.
Solving the multiple instance problem with axis-parallel rectangles.
Artificial Intelligence, 89(1-2):31–71, 1997.

K. Hempstalk, E. Frank, and I.H. Witten.
One-class classification by combining density and class probability estimation.
In *Proceedings of ECML PKDD*, pages 505–519. Springer, 2008.

V.N.and Severov D.A. Izraehl, Y.A.and Petrov.
Modelling of the transport and fallout of radionuclides from the accident at the chernobyl nuclear power plant.

In *International Symposium on Environmental Contamination Following a Major Nuclear Accident*. IAEA, 1989.

P. Langley.
Selection of relevant features in machine learning, 1994.

B. Lauritzen and T. Mikkelsen.
A probabilistic dispersion model applied to the long-range transport of radionuclides from the Chernobyl accident.
Atmospheric environment, 33(20):3271–3279, 1999.

T.J. Lyons and W.D. Scott.
Principles of Air Pollution Meterorology.
Belhaven Press, London, England, 1990.

MC MacCracken, DJ Wuebbles, JJ Walton, WH Duewer, and KE Grant.
The Livermore regional air quality model: I. Concept and development.
Journal of Applied Meteorology, 17(3):254–272, 1978.

AP Oliveira, J. Soares, T. Tirabassi, and U. Rizza.
A surface energy-budget model coupled with a skewed puff model for investigating the dispersion of radionuclides in a sub-tropical area of Brazil.
Nuovo cimento della società italiana di fisica. C, 21(6):631–646, 1998.

JR Simmonds, G. Lawson, and A. Mayall.
A methodology for assessing the radiological consequences of routine releases of radionuclides to the environment.
In *Proceedings of the Topical Meeting on Environmental Transport and Dosimetry: Charleston, South Carolina, September 1-3, 1993*, page 173. Amer Nuclear Society, 1993.

J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani.
1-norm support vector machines.
In *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*, pages 49–56. The MIT Press, 2004.