**Prediction methods**

In order to predict pollutants levels (quantitative variable) we can use the methods below :

1. Linear regression (simple or multiple)
2. Regression trees
3. Principal Component Analysis
4. Linear regression and PCA combined together - Principal Component regression.
5. Neural Network algorithm.

multilayered perceptron feed forward artificial neural network.

**Needed data**

We are going to need explanatory variable that could be :

* The area
* The date (maybe by separating the day of the week and the month)
* The time of the day
* The wind speed/direction (if relevant)
* The temperature (if relevant)
* The pollutants eg: carbon monoxide, ozone, particle pollutants, sulfur dioxide, nitrogen dioxide, methane, non methane hydrocarbon and total hydrocarbon.

For weather related data we’re going to scrape live data (is there existing API we can use ?)

**Softwares**

* Tanagra
* Sipina
* R-project
* Weka
* Orange
* JMP

**Principal Component Analysis**:

Dimension of a huge data set can be trimmed down using PCA. It is done in iterations, the more dimensions, the more number of iterations.

**Principal Component Regression:**

Variables are transformed into a set of orthogonal or uncorelated values.

Handle the problem of multicollinearity and produce stable and meaningful estimates for regression coefficients.

**Multiple Linear Regression Model:**

Model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data.  
 ***y*i =** **0 +** **1*x*i1 +** **2*x*i2 + ...** **p*x*ip +** **i for *i* = 1,2, ... *n*.**

where n is the number of observations.

**Neural Network Model:**

Pattern recognition.

Use input and generate output by providing a strength connection value to the input.

Back propagation could be used to determine if the strength connection value set for the inputs are right.

------------

Tools that work with PostgresSQL:

* MADLib library <http://madlib.net/>

<http://things-about-r.tumblr.com/post/98652993554/deep-down-below-using-in-database-analytics-from>

* R

<http://www.r-bloggers.com/getting-started-with-postgresql-in-r/>

* Weka (Java library)

<https://weka.wikispaces.com/How+do+I+connect+to+a+database%3F>

* Orange (Python)
* JMP (SAS solution, works with postgresql but only 30 days trial)