Middle-East Journal of Scientific Research 24 (9): 2727-2733, 2016
ISSN 1990-9233
© IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.09.23720

Rain Flow Model Based Air Pollution Estimation for Multi Attribute Environment Planning Using Data Mining

[1]A. Vinayagam, [2]C. Kavitha and [2]K. Thangadurai

[1]Department of Computer Science, Government Arts College (Autonomous),
Karur - 639 005, Tamilnadu, India
[2]Department of Computer Science, Thiruvalluvar Govt Arts College,
Rasipuram, Tamil Nadu, India

Abstract: The growth of population and the growth of human occupied region in the whole geographic of the world have many impacts in the environmental factors. The air pollution is one among them which has greater impact in various problems and has great impact in rain flow model. The environmental planning of any country has in great deal of estimating the air pollution incurred in each time window. Being reviewed various factors of environmental conditions, we propose a rain flow model which is generated based on air pollution of any region and computed using multi attribute of air pollution like the population, vehicle movement, agriculture lands, geographic regions, forest regions and the number of cyclones crossed, number of rainy seasons and amount of rain obtained at each time window. Using all these factors the model estimates the pollution of any region and estimates the probable rain could occur in that region at the next time window. This helps the environmental planning commission to decide on the focus to be given in increasing the forest regions, cultivating more plants in the forest and many more.

Key words: Air Pollution · Rain Flow Model · Global Warming · Data Mining · Pollution Estimation

INTRODUCTION

The growth of population has increased the usage of vehicles as well as the growth of human usage lands. This indirectly reduces the amount of agriculture lands and forest regions. For example, due to the increase of population, the human destroys the agriculture lands and convert them into boarding lands and they construct the houses in the regions. Also the increase of population destroys the region of the forest by cutting the plants [1].

The problem is the human is not cultivating the new plants in the forest where there are many plants gets destroyed due to the fire introduced by different plants collide. This collision destroyed more region of plants and increases the global warming. Also the usage of vehicles and the increasing population also, increase the global warming conditions. The global warming conditions increase the temperature enormously and affect the rainy conditions also.

Not only the scarcity of plants but also the air pollution has also had great impact in the rain scarcity, because the ozone layer gets damaged at all the time because of the air pollution introduced by the people of world. The pollution introduced by the human can be estimated by various factors like, the number of vehicles in any geographic region, the fuel consumption of the region, traffic parameters, number of industries, number of industries has fixed filters, geographic region, amount of forest region, size of agric lands and many more. By keep tracking of these parameters in different time window, the rain flow can be estimated and air pollution can be estimated at different time window.

Once the air pollution can be estimated then the planning commission can estimate the amount of plants to be cultivated at each region and how the pollution can be controlled in each region to increase the rain flow of each region, which ultimately reduces the global warming condition. These all can be done by using data mining

Corresponding Author: A. Vinayagam, Department of Computer Science, Government Arts College (Autonomous), Karur - 639 005, Tamilnadu, India.

techniques. Data mining is a computing approach to extract information required from large set of data. In air pollution estimation, the data mining ideology could be used where to mine information from large set of traffic data and to infer useful information about traffic pattern and how it affects the environment conditions. Similarly from other factors we could identify and extract useful information's from the large set of data.

Related Works: There are few works which has been proposed in this area and we discuss about them here.

iMAP: Indirect measurement of air pollution with cellphones [1], introduce the cell phone-based indirect sensing problem. While participatory sensing aims at monitoring of a phenomenon by deploying a dense set of sensors carried by individuals, our indirect sensing problem aims at inferring the manifestations of a sparsely monitored phenomenon on the individuals. The main advantage of the indirect sensing method is that, by making use of existing exposure modeling and estimation methods, it provides a more feasible alternative to direct sensing. Collection of time-location logs using the cellphones plays a major role in our indirect sensing method, while direct sensing at the cellphones is unneeded.

SO2 classification for air quality levels estimation using artificial intelligent techniques [2], presents a new methodology to detect and classify SO_2 concentration according to the air quality level. In this classification, meteorological variables are analyzed to make a classification decision. The method consists of three steps. In first step, we group using a SOM neural networks the pollutant concentration in two classes, these classes are noise data and validated data. In second step, we create a representative feature vector using the information contingency levels that we know a priori. In third step, a new SOM neural network is trained with the representative feature vector built in second step and then the pollutant concentrations and meteorological variables (validated data) are self-organized in fourth classes according to contingency levels. Finally, we obtained the air quality level.

Dynamic estimation of air pollution [3], propose advection-diffusion model of air pollution over an urban area. The region is subdivided into a grid and the three-dimensional partial differential equation of the pollution concentration is reduced to a linear vector difference equation. Along with this discrete equation, a stochastic model of air pollution is considered and pollution concentrations over the area are estimated from observed data generated by a few monitor points.

In Estimation of exhaust emissions of marine traffic using Automatic Identification System data [4], an Automatic Identification System (AIS) receiver is used to obtain ship data. AIS recognize a vessel's Maritime Mobil Maritime Identify (MMSI), speed of ship, initial position of ship and ship type. This data is used to evaluate the marine traffic density in the Madura Strait area. Information from ship databases and AIS data are combined for retrieving gross tonnage (GT) information, which is then used to estimate the ship's air pollution emissions. Air pollution estimates also consider the ship's operation modes such as berthing, maneuvering and hostelling.

Air pollution data classification by SOM Neural Network [5], presents a Self-Organizing Maps (SOM) Neural Network application to classify pollution data and automatism the air pollution level determination for Sculpture Dioxide (SO2) in Salamanca. Meteorological parameters are well known to be important factors contributing to air quality estimation and prediction. In order to observe the behavior and clarify the influence of wind parameters on the SO2 concentrations a SOM Neural Network have been implemented along a year. The main advantages of the SOM are that it allows integrating data from different sensors and providing readily interpretation results. Especially, it is powerful mapping and classification tool, which others information in an easier way and facilitates the task of establishing an order of priority between the distinguished groups of concentrations depending on their need for further research or remediation actions in subsequent management steps.

Influence of Air Pollution on Cardiovascular Diseases Prevalence in Developing Countries: An Eco-Social Model [6] proposes the use of eco-social analysis supported by Geographical Information System (GIS) to derive more accurate results. Local/Urban-Scale PM10 Concentration Estimation from TM Imagery [7] offers a unique opportunity to estimate air quality that is critically important for the management and surveillance of air quality in some cities of China, which have experienced elevated concentration of air pollution but lack adequate spatial-temporal coverage of air pollution monitoring. A local/urban-scale PM10 concentration estimation method is developed with 30m resolution TM imagery. The method can be used for local/urban-scale imagery under a variety of atmospheric and surface conditions.

Development of a real-time on-road emissions estimation and monitoring system [8], estimate and monitor operational on-road emissions with high accuracy and resolution in real time. The two sets of critical information for emission estimation, vehicle mix and vehicle activity, are directly generated from traffic detection using inductive vehicle signature technology. An initial implementation on a section of the I-405 freeway at Irvine, California is demonstrated. With more widespread deployment, the system can be used to perform before-and-after evaluation of certain mitigation strategies, to develop time sensitive optimal traffic control strategies with the purpose to control emissions and to provide high fidelity greenhouse gas and air quality information to policymakers, researchers and the general public.

Estimating NH3 emissions from agricultural fertilizer application in China using the bi-directional CMAQ model coupled to an agro-ecosystem model [9], estimate the NH3 emission from the agricultural fertilizer application in China online using an agricultural fertilizer modeling system coupling a regional air quality model (the Community Multi-Scale Air Quality model, CMAQ) and an agro-ecosystem model (the Environmental Policy Integrated Climate model, EPIC), which improves the spatial and temporal resolution of NH3 emission from this sector. Cropland area data of 14 crops from 2710 counties and the Moderate Resolution Imaging Spectroradiometer (MODIS) land use data are combined to determine the crop distribution. The fertilizer application rate and method for different crop are collected at provincial or agriculture-regional level. The EPIC outputs of daily fertilizer application and soil characteristics are inputed into the CMAQ model and the hourly NH3 emission are calculated online with CMAQ running. The estimated agricultural fertilizer NH3 emission in this study is about 3 Tg in 2011. The regions with the highest modeled emission rates are located in the North China Plain. Seasonally, the peak ammonia emissions occur from April to July.Compared with previous researches, this method considers more influencing factors, such as meteorological fields, soil and the fertilizer application and provides improved NH3 emission with higher spatial and temporal resolution.

Atmospheric concentration measurements are used to adjust the daily to monthly budget of fossil fuel CO2 emissions of the Paris urban area from the prior estimates established by the Airparif local air quality agency. Five atmospheric monitoring sites are available, including one at the top of the Eiffel Tower. The atmospheric inversion is based on a Bayesian approach and relies on an atmospheric transport model with a spatial resolution of 2 km with boundary conditions from a global coarse grid transport model. The inversion adjusts prior knowledge about the anthropogenic and biogenic CO2 fluxes from the Airparif inventory and an ecosystem model, respectively[10], with corrections at a temporal resolution of 6 h, while keeping the spatial distribution from the emission inventory. These corrections are based on assumptions regarding the temporal autocorrelation of prior emissions uncertainties within the daily cycle and from day to day in [11].

Evaluation of the high resolution WRF-Chem air quality forecast and its comparison with statistical ozone predictions [12], propose an evaluation of the air quality forecasting system has been performed for summer 2013. In the case of ozone (O3) daily maxima the first day and second day model predictions have been also compared to the operational statistical O3 forecast and to persistence. Results of discrete and categorical evaluations show that the WRF-Chem based forecasting system is able to produce reliable forecasts, which depending on monitoring site and the evaluation measure applied can outperform the statistical model. For example, correlation coefficient shows the highest skill for WRF-Chem model O3 predictions, confirming the significance of the non-linear processes taken into account in an on-line coupled Eulerian model. For some stations and areas biases were relatively high due to highly complex terrain and unresolved local meteorological and emission dynamics, which contributed to somewhat lower WRF-Chem skill obtained in categorical model evaluations. Applying a bias-correction could further improve WRF-Chem model forecasting skill in these cases.

All these approaches are focused only towards real time data and misses the earlier history to be visited before estimating the air pollution. We propose such a model to predict the upcoming air pollution with the help of time variant data and data mining techniques [13].

Proposed Method: The proposed rain flow model based air pollution estimation has various components namely preprocessing, Time variant environment pattern generation, Air pollution estimation and Rain flow estimation. We discuss each of the functional components in detail in this section.

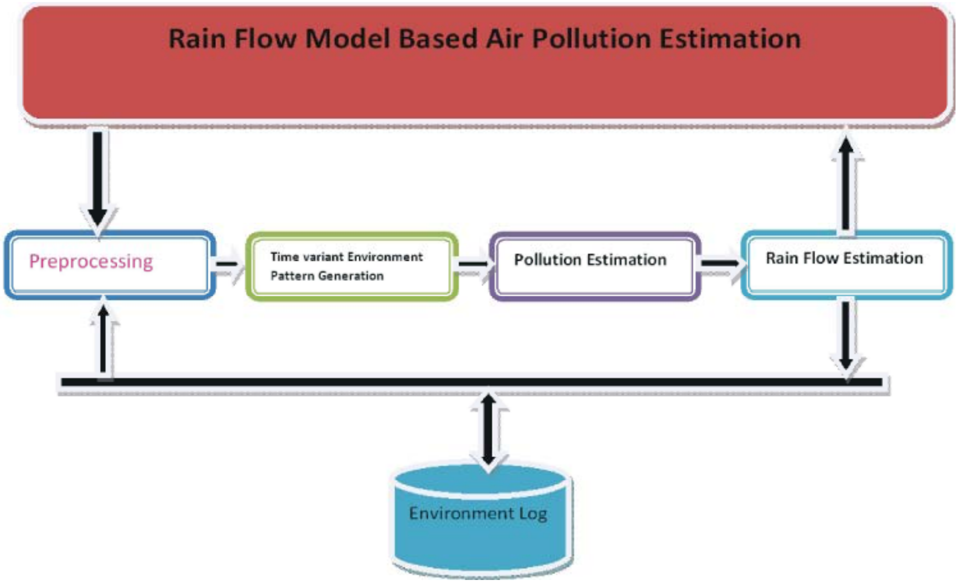

Rain Flow Model Based Air Pollution Estimation
Preprocessing
Time variant Environment Pattern Generation
Pollution Estimation
Rain Flow Estimation
Environment Log

Fig. 1: Proposed System Architecture.

The Figure 1 shows the architecture of the proposed system and its functional components. We discuss the functionality of the each component in detail in this section.

Preprocessing: Preprocessing is the process of identifying the exact data from the environment log and removing the incomplete one. Also the method converts the log into processing form which will help to generate patterns in the next stage easier.

```
Pseudo Code:
Input: Environment Log El.
Output: Preprocessed Log Pl.
Step1: start
Step2: Identify the number of distinct feature of the log.
        Log Feature LF = $\sum_{i=1}^{size(El)} \sum Attribute(El(i)) \exists LF$
Step3: for each log Li from El
        If Li $\in \sum Attr(LF) then$
                Add log to Pl.
        else
                remove log.
        end
    end
Step4: stop.
```

Time variant Environment Pattern Generation: At this stage, the preprocessed log is split into number of time window and for each window we compute various measures and convert them into feature vector. The features of the log is retrieved or extracted from the log and the extracted feature will be used to form the time variant environment pattern. For each of the feature considered, we compute the mean value to generate the environment pattern.

```
Algorithm:
Input: Preprocessed Log Pl.
Output: Environment Pattern Ep.
Step1: start
Step2: for each time window Two
        Identify logs belong to the time window tw.
        Tl = $\sum_{i=1}^{size(Pl)} \sum Pl(i).Time == Tw$
        Identify the zone value Zone.
        compute Geographic area Garea= size(GR).
        Compute Forest Area FA = GArea-(Aarea+HArea).
        Compute Harea = $\sum squareMeters(Houses)$
        Compute AgriArea AA = $\sum squareMeters(Agri\ Lands)$
        Compute Population Pp = $\sum Peoples \in Region$
        Compute Vehicles Nv = $\sum Vehicles \in Region$
        Compute Fuel consumption Fc = NV×μ.
        Compute Industry factor NI = $\sum Industries \in Region$
        Compute zone temperature Zt = Temp $\in Region$
        Compute number of rain falls NF = $\sum RF \in Region$
        Compute number of cyclones NCY = $\sum Cyclones \in Region$
        Compute average centimeter fall Ac = $\frac{\sum centimeters\ of\ rain}{number\ of\ times}$

    Generate pattern Tp = {Zone, GA, FA, HA, AA, Pp, NV, FC, NI, ZT, NF, NCY, AC}.
        Add pattern to pattern set Ep = $\sum(Pi \in Ep) + Tp$
    End
Step3: stop.
```

Pollution Estimation: The pollution at any region at any point of time is computed from the environment pattern being generated. The proposed method estimates the pollution using the probabilistic model. Because the pollution occurred in the previous time window has records but there is no record will be available for the current status and future happenings. So that the pollution will be estimated as follows.

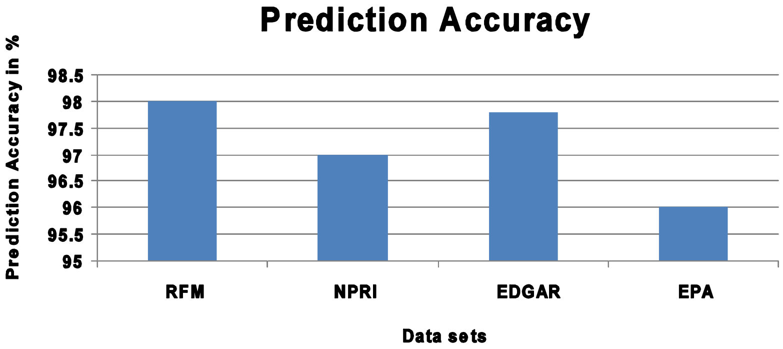
```
Input: Pollution Pattern PoP.
Output: Current Pollution and Future Value.
Step1: Split time zone into N.
Step2: for each time zone
        identify patterns $PoP_T = \int_{Tstart}^{Tend} PoP(p) \in p(Time)$
    end.
Step3: for each time zone
        Compute average pollution occurrence Apc as follows.

        Compute geographic impact factor Gif.

    GIR=($\sum_{1}^{N}(PoP(Gr) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

        Compute forest impact factor Fif.

    FIR = ($\sum_{1}^{N} PoP(Fr)$/N) ×. ($\sum_{1}^{N}(PoP(Pollution) \in Tw)/N$.

        Compute Agriculture impact factor Aif.

    AIR = ($\sum_{1}^{N}(PoP(Ar) \in Tw)$/N) ×. ($\sum_{1}^{N} PoP(Pollution)/N$.

        Compute Population impact factor Pig.

    PIR = ($\sum_{1}^{N}(PoP(po) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

        Compute vehicle impact factor Vif.

    VIF = ($\sum_{1}^{N}(PoP(Veh) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

        Compute Fuel consumption impact factor Fcif.

    FCIF= ($\sum_{1}^{N}(PoP(Fc) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

        Compute Industry impact factor IF

    IF = ($\sum_{1}^{N}(PoP(In) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

        compute zone temperature impact factor ZTF.

    ZTF = ($\sum_{1}^{N}(PoP(Zt) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

        compute rain fall factor ZRF.

    ZRF = ($\sum_{1}^{N}(PoP(Rf) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

    compute cyclone factor ZCF.

    ZCF = ($\sum_{1}^{N}(PoP(Cyclone) \in Tw)$/ N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

    compute centimeter factor ZCCF.

    ZCCF = ($\sum_{1}^{N}(PoP(centimeter) \in Tw)$/N)×.($\sum_{1}^{N} PoP(Pollution)/N$.

    Compute pollution p =(Σ Gif + Fif + Aif + Pif + Vif + Fif + If + ZTf + ZRF + ZCF + ZCCF)/11
    end.

Step4: compute current pollution CP.

    Current pollution Cp=(Σ Gr + Fr + Ar + Pr + Vr + Fcr + Ir + Zr)/8

Step5: Future Pollution Ratio Fpr.
        Compute standard deviation on each of the time window air pollution.

        Future Ratio Fr = $\sum_{i=1}^{size(Tw)} Stddev(Pollution\ Ratio)$

Step6: Stop.
```

Rain Flow Estimation: The rain flow estimation is performed based on the estimated air pollution. Based on the air pollution computed at each time window the method computes the rain flow estimation which specifies the possibility of rain could occur at the next time window. This will helps the environment planning commission to generate plans accordingly.

```
Algorithm:
Input: Time variant environment pattern set Tps, Pollution set Ps.
Output: Rain Flow Estimation RFS.
Step1: start
Step2: for each time window tw
        compute rain factor Rf = $\sum_{i=1}^{size(Tps)} \frac{\sum ZRF+ZCF+ZCCF}{size(Tps(tw))}$
        Compute rain probability Rpb= Tw (Pollution) ×Rf.
    End
Step3: Compute standard deviation Std-Rf = Stddev(Rpb).
Step4: Compute Future rain probability Frp = Ø×Log(Std-Rf).
Step5: stop.
```

RESULTS AND DISCUSSION

The proposed method has been evaluated with the data set available at. The proposed method has been tested for its efficiency in prediction and accuracy. We have used various size of data set to evaluate the performance of the proposed approach. We have used NERC data set, which is provided by open air project, Canada and NPRI data set of Canada.



Graph 1: Shows the prediction accuracy achieved.

The Graph1 shows the value of prediction accuracy achieved by the proposed method with different data sets. It is clear that the proposed method has produced higher efficient results with all the data sets available.

CONCLUSION

This paper presented a rain flow model based air pollution estimation using data mining. The pattern based data mining approach has been used for pollution estimation. The proposed method used all the factors of air pollution, so that to increase the performance of the pollution estimation. Unlike other methods we have used histories of air pollution and dynamic details also. The proposed method has produced efficient results.

REFERENCES

1. Demirbas, M., 2009. iMAP: Indirect measurement of air pollution with cellphones, IEEE Conference on Pervasive computing and communications, pp: 1-6.
2. Cortina-Januchs, M.G., 2008. SO_2 classification for air quality levels estimation using artificial intelligent techniques, Multiconference on Electronics and photonics, 158-162.
3. Desalu, 2003. Dynamic estimation of air pollution, IEEE Transaction on Automatic Control, 19(6): 904-910.
4. Pitana, 2010. Estimation of exhaust emissions of marine traffic using Automatic Identification System data, IEEE OCEANS, pp: 1-6.
5. Barron-Adome, 2012. Air pollution data classification by SOM Neural Network, World Automation Congress, pp: 1-5.
6. Olayanju, L.O., 2011. Influence of Air Pollution on Cardiovascular Diseases Prevalence in Developing Countries: An Eco-Social Model, Developments in E-System Engineering, pp: 50-55.
7. Weiwei Song, 2010. Local/Urban-Scale PM10 Concentration Estimation from TM Imagery, IEEE conference on Bioinformatics and Biomedical Engineering (iCBBE).

8. Hang Liu, 2011. Development of real-time on-road emissions estimation and monitoring system, IEEE conference on Intelligent Transport System, pp: 1821-1826.
9. Fu, X., S.X.Wang, L.M. Ran, J.E. Pleim, E. Cooter, J.O. Bash, V. Benson and J. Hao, 2015. M.: Estimating NH3 emissions from agricultural fertilizer application in China using the bi-directional CMAQ model coupled to an agro-ecosystem model, Atmos. Chem. Phys. Discuss., 15: 745-778.
10. Cheng, Z., S. Wang, X. Fu, J.G. Watson, J. Jiang, Q. Fu, C. Chen, B. Xu, J. Yu, J. C. Chow and J. Hao, 2014. Impact of biomass burning on haze pollution in the Yangtze River delta, Atmos. Chem. Phys., 14: 4573-4585.
11. Bréon, F.M., G. Broquet, V. Puygrenier, F. Chevallier, I. Xueref-Remy, M. Ramonet, E. Dieudonné, M. Lopez, M. Schmidt, O. Perrussel and P. Ciais, 2015. An attempt at estimating Paris area CO2 emissions from atmospheric concentration measurements, Atmos. Chem. Phys., 15: 1707-1724.
12. Žabkar, R., L. Honzak and G. Skok, 2015. For doi: 10.5194/acp-15-1707-2015.
13. Rarely, J. Rakovec, A. Ceglar and N. Zagar, XXXX. Evaluation of the high resolution WRF-Chem air quality forecast and its comparison with statistical ozone predictions, Geosci. Model Dev. Discuss., 8: 1029-1075.