

PREDICTING THE EFFECT OF FINE PARTICULATE MATTER (PM_{2.5}) ON AN ECOSYSTEM INCLUDING CLIMATE, PLANTS AND HUMAN HEALTH USING MACHINE LEARNING METHODS

Subhadra Rajpoot, Prakarsh Kaushik, Devang Pratapsingh, Brij Kishore Tiwari

Depart. Of Applied Sciences Amity School Of Engineering and Technology Greater Noida, India
Student of Computer Science and Engineering Amity School Of Engineering and Technology Greater Noida, India

Department of applied science and Humanities, G.L. Bajaj Institute of Technology and management
Gr. Noida (U.P.) Pin-201306

ABSTRACT

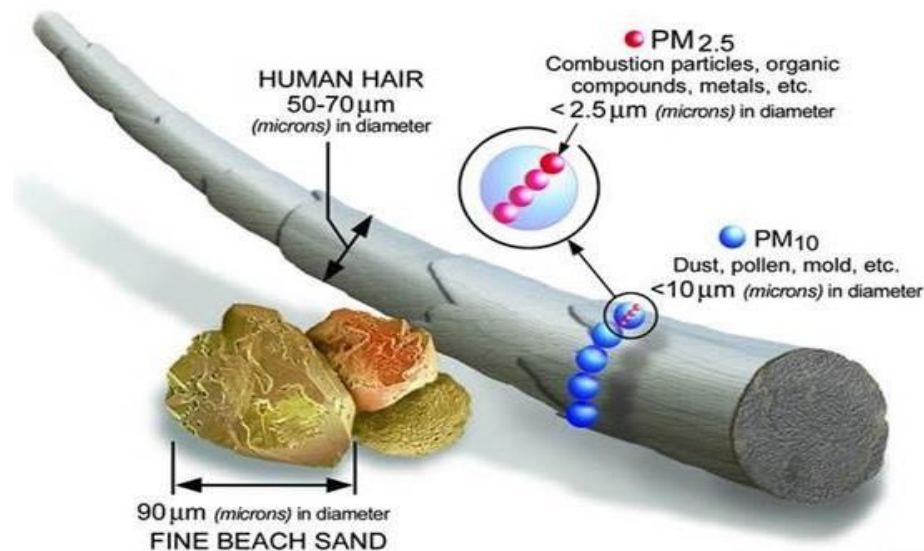
Particulate matter (PM) is an unexpected mixture of solid and fluid molecules suspended in the atmosphere. These particles consist regular basis of a mixture of inorganic and organic synthetic substances, including carbon, sulphates, nitrates, metals and acids. Research advises that molecular size is a key factor that influences the respiratory tract and affects humans. Large and rough particles are deposited only in the nasal passages; however, tiny and ultra-fine particles, can penetrate deep areas of the lung. Very fine and minute particles are present in much more noteworthy figures and have a much more prominent surface area than particles bigger in size of similar weight, and are usually rather more hazardous. Some countries contribute to ambient particulate matter contaminants in other countries. Media coverage has been extended to particulate matter in recent years, as a result of the discovery of thousands of young people kicking a bucket of particulate matter posed well-being impacts. It covers both the current time and the future. The particulate matter problem does not only illustrate well-being impacts; it also plays a role in the influence of a green house and an abnormal shift in climate as a result of its participation to cloud arrangement. The study disclosed in this paper is crucial for a much broader exploratory study on airborne particulate matter. These particulates have now become a subject of real concern in the global environment, as they may not only have generous negative effects for human well-being, but they may also have an effect on rainfall pattern, the world's ecosystem and well-being of flora and fauna. In addition, the presence of these particles on vegetated surfaces can adversely affect the life of plants and creatures. The dimension, composition of the substances and the cause of these particles are tremendously changing. The impacts of these particles have been studied, evaluated and discussed in this paper with their corresponding natural influence.

Keywords—climate, plants and human health.

INTRODUCTION

Particulate matter is also referred to as atmospheric aerosol particles, particulate or suspended particulate matter (SPM). (SPM). At the most basic level, particulate matter may be a mutual term that encompasses a mixture of many small solid and liquid particles suspended in the Earth's atmosphere. These particles often include dust, fly ash, soot, smoke, aerosols, fumes, mists and condensing vapors that will be suspended in the atmosphere for a significant period of time. For the last two years, particulate matter has gained substantial public coverage, primarily due to heightened morality from particulate-related health diseases. Ecological degradation has also been related to airborne particulate matter, especially in developed nations where the accumulation of these particulate matter is very high as a result of construction/industrial developments. The harm resulting due to these particulates in developed countries is projected at billions of money. Particulate matters have also attracted interest from the scientific community because of its obvious role/contribution in weather pattern and global climate change are now a matter of in-depth study.

The particulate matter origins include wind-blown dust and forest fires. Secondary PM sources specifically inject air pollutants into the environment that shape or lead to the production of particulate matter. These contaminants are thus essentially called precursors to the formation of PM. These secondary contaminants are SO_x, NO_x, VOCs and NH₃. There are many different types of particulate matter, which can be distinguished into different groups, depending on their size; Complete suspended particles, including all particles of any size; PM₁₀ particles of 10 μm (10 microns) in diameter; PM_{2.5} particles of less than 2.5 μm in diameter. Finespores, which are smaller than 2.5 μm in diameter, are the most harmful in the group of particulates.



I. DEPOSITION PROCESSES

Particulate matter can gather or store themselves on strong surfaces, through two sub-measures: dry and wet deposition. The past incorporates affidavits by impaction, capture, gravitational sedimentation, choppiness and other supportive processes like thermophoresis, electrophoresis and so forth while in the last condition, barometrical hydrometeors (downpour drops, snow and so on) search airborne particles. This basically implies wet deposition is gravitational, Brownian or potentially tempestuous coagulation with water beads. These statements can initiate incredible changes in the environment. ensuing segment examines the differed impacts of particulate on plants, atmosphere, human well-being and environment by and large

Barometrical particulate matter (PM) deposition which includes both dry and wet cycles is a significant method for controlling air contamination. To research the attributes of dry and wet deposition in wetlands, PM concentrations and meteorological conditions were monitored during summer at heights of 1.5m, 6m and 10m above ground level in different locations in order to assess the efficiency of PM_{2.5}. It was likewise noticed that rain fall power and PM breadth impacted wet deposition effectiveness. Dry deposition (63%) was more tilted towards removing PM₁₀ than was the case for wet deposition (37%). In terms of PM_{2.5} removal, wet deposition (92%) was found to be more efficient

The regular event of particulate matter (PM) contamination has prompted numerous issues. The outflow of PM is perhaps the main variable influencing atmosphere and well-being. Environmental testimony is a significant method for controlling air contamination. Climatic PM testimony includes both dry and wet cycles. Dry deposition alludes to the affidavits of particles or gases from the air through the immediate conveyance of mass to the surface (for example through non-precipitation). Then again, wet cycles are frequently alluded to as downpour or snow searching, with downpour rummaging PM being by and large named 'rainout particles' (filling in as cloud-buildup cores or going through catch by cloud water) and as 'waste of time' (for example evacuation of beneath cloud particles by rain drops as they fall). Studies have affirmed that dry deposition has the ability to eliminate PM and that the cycle of dry deposition is affected by spatial vacillations, surface-type contrasts, transient changes, diurnal varieties and meteorological conditions. Wet deposition is likewise a significant system for decreasing air contamination.

II. METHODS AND METHODOLOGIES

A. Machine Learning Algorithms

The detailed specifics of how these machine learning algorithms work is described. There are different types of algorithms in Machine

Learning. We often tend to apply all of these algorithms without thinking as to when to apply what. When we know the differences in between these algorithms, it becomes easier for us to optimize their application and know in which scenario we use what. There are certain parameters on which these algorithms differ. Understanding these parameters and their application can help us attain better accuracy and minimize the chances of error in our predictions.

The Gradient boosting machine (GBM) and Random Forest (RF) are both decision tree machine learning algorithms. The GBM is a boosting decision tree method. This suggests that weak learners are created and therefore the residuals from these models are used to create stronger learners, with previous models essentially “boosting” subsequent models and improving the predictions.

In order to achieve higher accuracy, RF decides to create a large number of them based on bagging. The basic idea is to sample the data over and over and for each sample train a new classifier. Different classifiers overfit the data in a different way, and through voting those differences are averaged out. The process of averaging or combining the results of different decision trees helps to overcome the problem of overfitting. Random forest also has less variance than a single decision tree. It means that it works correctly for a large range of data items than single decision trees. During a random forest, an outsized number of decision trees are constructed and predictions from these individual trees are averaged together to output.

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature. Neural networks can adapt to changing input; so, the network generates the best possible result without needing to redesign the output criteria. The neural network takes the input variables, very similar to a neuron responds to stimuli, processes them through various combinations and weights, and generates predictions. Neural networks are retrained and taught just like a child's developing brain is trained. They cannot be programmed directly for a particular task. They are retrained in such a manner so that they can adapt according to the changing input.

The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement, non-parametric, lazy learning, supervised machine learning algorithm that can be used to solve both classification and regression problems using feature similarity. Learning KNN machine learning algorithm is a great way to introduce yourself to machine learning and classification in general. The k-nearest neighbor algorithm relies on the idea that there's a proximal relationship between values. It calculates a weighted average from the closest designated “k” neighbors, so as to get prediction. Each machine learner was run separately on the info.

What is K- Nearest Neighbors? K- Nearest Neighbors is a –

- ⇒ Nonparametric as it does not make an assumption about the underlying data distribution pattern
- ⇒ Lazy algorithm as KNN does not have a training step. All data points will be used only at the time of prediction. With no training step, prediction step is costly.
- ⇒ Supervised machine learning algorithm as target variable is known
- ⇒ Used for both Classification and Regression
- ⇒ Uses feature similarity/nearest neighbors to predict the cluster that the new point will fall into.

B. Generalised Linear Models (GLMs)

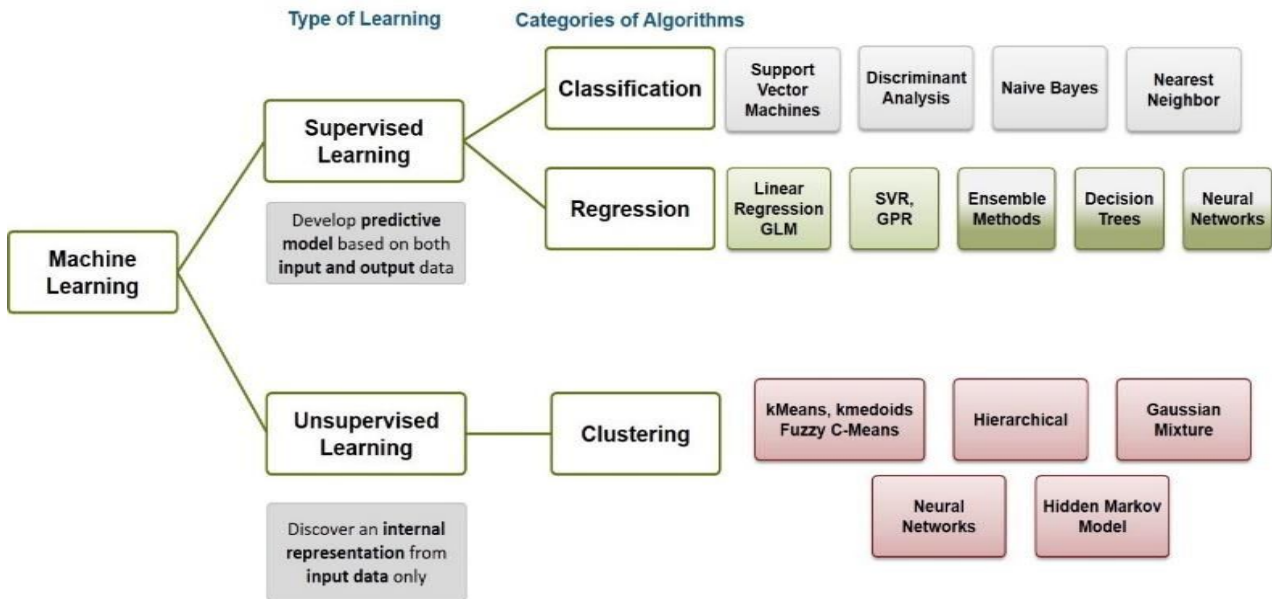
It explains how Linear regression and Logistic regression are a member of a much broader class of models. GLMs can be used to construct the models for regression and classification problems by using the type of distribution which best describes the data or labels given for training the model.

C. Generalised Additive Models (GAMs)

are semi-parametric extensions of GLMs, only making an assumption that the functions are additive and the components are smooth. GAMs have the ability to deal with highly non-linear and non-monotonic relationships between the response and explanatory variables. The individual machine learners were ensemble-averaged employing a GAM, including a smoothed function of the predictions from each individual learner, plus a smoothed function of latitude and longitude. The predictions from this GAM are the ensemble-averaged predictions. The smoothing

terms allow the weights given to every learner to vary with the pollution level, in case one learner performs better during a specific range of PM2.5

When to use GAMs



- When assumptions cannot be made on specific link function for error distribution
- Non-linearity in partial residual plots may suggest semi-parametric modeling
- Prior hypothesis or theory suggest non-linear or skewed relationship among variables
- Shape of predictor functions is determined by the data (Data speak for themselves!!)

i. INPUT VARIABLES

The number of input variables or features for a dataset is referred to as its dimensionality. Dimensionality reduction refers to techniques that reduce the number of input variables in a dataset. More input features often make a predictive modelling task more challenging to model, more generally referred to as the curse of dimensionality. High-dimensionality statistics and dimensionality reduction techniques are often used for data visualization. Nevertheless, these techniques can be used in applied machine learning to simplify a classification or regression dataset in order to better fit a predictive model.

The covariates utilized in training the models were population density (persons/km²), cloudiness (okta), atmospheric pressure (mBar/hPa), wind direction (°N), wind speed (m/s), dewpoint temperature (°C), temperature (°C), aerosol optical depth (AOD), land use type, distance to water (km), distance to Heathrow airport (m), inverse of the peak of the planetary physical phenomenon (m-1), normalized difference vegetation index (NDVI), traffic counts, distance to nearest major road (km), length of major road (km) in grid cell, number of bus stops in grid cell, distance to nearest bus stop (km), average building height (m), and number of buildings within the grid cell. We selected these variables, based on expert knowledge and data availability. Meteorological variables are known to affect the dispersion and transport of fine particulate. Land-use variables represent potential sources of PM2.5 and areas of upper concern. The time variables account for the seasonal variation in PM2.5 levels and therefore the trend over several years. As previously mentioned, AOD may be a key predictor of PM2.5, with higher levels of AOD indicating higher PM2.5 levels [18–22]. We also used cross-validated R² values within the GBM algorithm to work out whether additional variables would improve model performance choose whether to incorporate them or not.

ii. DATA SOURCES

Data ingestion is the process in which unstructured data is extracted from one or multiple sources and then prepared for training machine learning models. It's also time intensive, especially if done manually, and if we have large amounts of data from multiple sources. Automating this effort, frees up resources and ensures your models use the most recent and applicable data.

The meteorological variables were obtained from the Meteorological Office. Traffic counts were obtained from the Department of Transport. The average building height, number of buildings, distance to nearest major road from grid cell centroid, and length of major road in grid cell were calculated using data provided by the Ordnance Survey. There were missing AOD values due to cloud cover. We imputed these missing values using a random forest approach and land use and meteorological predictors. The random forest approach.

iii. PM2.5 DATA

Predicted PM2.5 was compared to measured PM2.5 to check the models and assess the accuracy of the varied methods used. Measured PM2.5 was obtained from different monitoring sites. Measured PM10 information was obtained at monitoring sites for fixed period. So as to predict PM2.5 at fixed sites with no PM2.5 measurements, but which measured PM10 and NOx, we used two methods: (1) a regression model and (2) a random forest approach. The predictions from the two methods were used as independent variables during a generalized additive model (GAM), so as to enhance the fit of the model and thus provide a greatly enhanced database of PM2.5 estimated concentrations. These values were then treated as our measured PM2.5, which we used for training and cross-validation.

iv. PREDICTIONS

"Prediction" refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be.

We used ten-fold cross-validation to see the robustness of our model. We divided the monitoring stations into ten groups. Each model was trained on data from ninety percent of the monitors and predicted within the held-out one-tenth. This process was repeated ten times to completely recreate the measured dataset from the portion of the info during which training didn't occur. We then checked out the correlation of the anticipated PM2.5 with the measured PM2.5. So as to see the model's ability to capture spatial variation, we compared annual average predicted PM2.5 to the measured annual average PM2.5 at monitoring sites, as seen within the equation below:

$$\text{Annual Measured PM2.5}_{ij} = \beta_0 + \beta_1 \text{Annual Predicted PM2.5}_{ij} \quad (1)$$

where i is the monitoring site and j is the year. In order to look at the temporal accuracy, we looked at the difference between predicted and measured PM2.5 levels and their annual averages, as seen in the equation below:

$$\begin{aligned} & \text{Daily Measured PM2.5}_{ij} - \text{Annual Measured PM2.5}_{ij} \\ &= \beta_0 + \beta_1 (\text{Daily Predicted PM2.5}_{ij} - \text{Annual Predicted PM2.5}_{ij}) \quad (2) \end{aligned}$$

We chose our final prediction model supported the general, spatial, and temporal adjusted R^2 values. To assess the linearity of the connection between predicted and measured PM2.5, we regressed

the final predictions against the measurements for both the spatial and temporal component, using a penalized spline, which chooses the degree of nonlinearity supported by the restricted maximum

likelihood. The spatial component was modeled using the subsequent equation:

$$\text{Annual Measured PM2.5}_{ij} = s(\text{Annual Predicted PM2.5}_{ij}) + \epsilon_{ij} \quad (3)$$

In this equation, i is the monitoring site and j is the year, and s is a smoothing function.

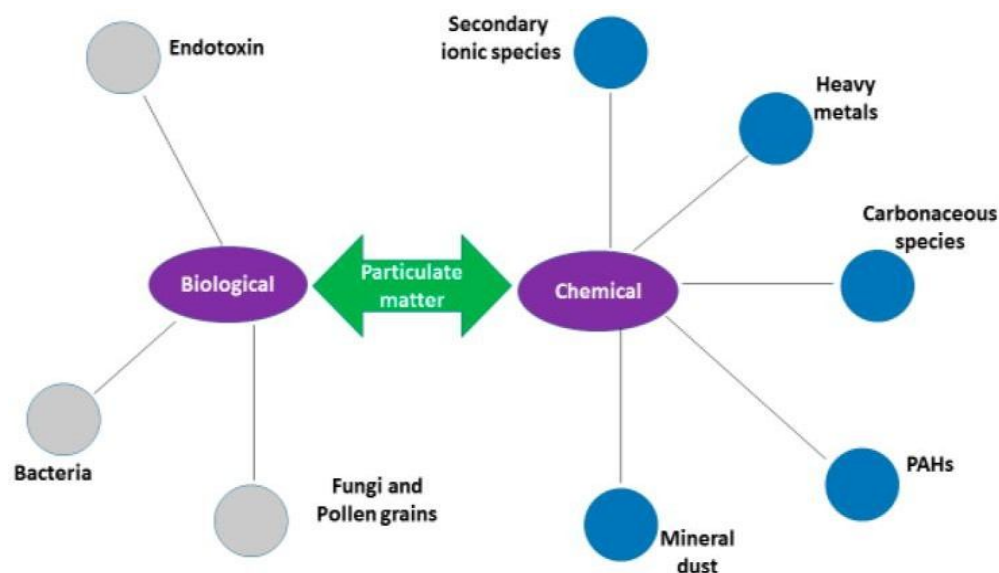
$$\begin{aligned} & \text{Daily Measured PM2.5}_{ij} - \text{Annual Measured PM2.5}_{ij} \\ &= s(\text{Daily Predicted PM2.5}_{ij} - \text{Annual Predicted PM2.5}_{ij}) + \epsilon_{ij} \quad (4) \end{aligned}$$

III. CLIMATE EFFECTS

The change inside the world's atmosphere is driven by specific irritations inside the world's energy balance. These both are on the whole referenced as "atmosphere driving", and are the subject of significant exploration for anticipating future change inside the world's atmosphere. Varieties in atmosphere driving are characterized or controlled by actual effects on the climate like orbital and pivotal changes likewise as by the presence of specific specialists like ozone harming substances and vaporized particles, which are fit for adjusting the world's energy balance. This progressively achieves changes inside the world's worldwide temperature and along these lines changes inside the atmosphere. The commitment of particulate during this change is very generous. Particulate can influence the atmosphere either straight forwardly or example through the technique for dissipating and assimilation of the radiation, or by implication for example through the

development of cloud buildup cores (CCN). The immediate commitment of particulate is because of the presence of sulfate pressurized canned products, fuel soots and discharges from biomass copying within the earth's atmosphere.

Wind speed, blending stature, and relative stickiness are the meteorological factors accepted to generally impact PM focuses. Stale conditions are thought to associate with high PM fixations, as they permit particulates to collect approach the world's surface. Albeit high wind rates can expand ventilation, they are ordinarily related with high PM fixations since they permit the resuspension of particles starting from the earliest stage, well as long-range transport of particulates between areas. High PM focuses are typically connected with dry conditions because of expanded potential to resuspension of residue, soil, and different particles.

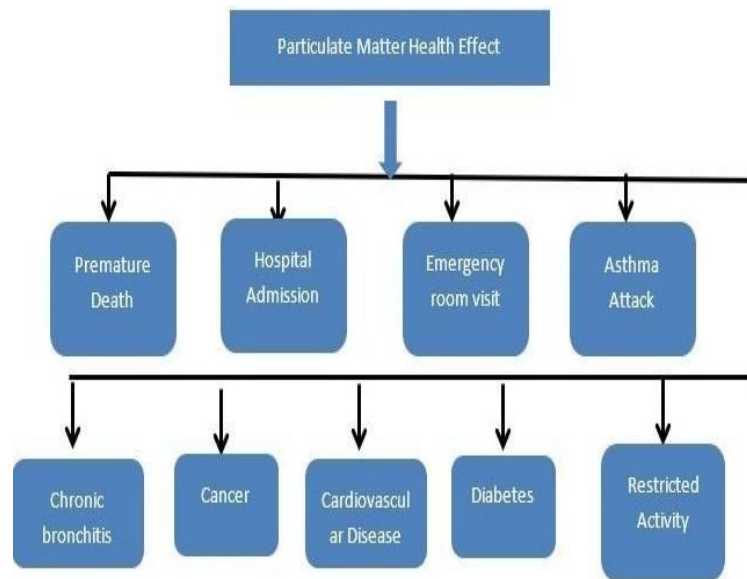


A. Scattering and absorption of light due to aerosol particles

The dispersing of daylight might be an overall actual cycle and is characterized on the grounds that the reallocation of daylight in non equal ways when it associates with little particles of issue. Light dissipating of airborne particles can happen because of the presence of a decent kind of vaporized sorts, among which the carbon and sulfate containing particles is accepted to be the chief proficient. Dispersing on account of airborne mode (particles whose breadth are of an equal request in light of the fact that the frequency of the occurrence light) delivers a negative driving (cooling impact) on the grounds that the dissipated light from these particles falls prevalently into the locale of Mie dissipating which is portrayed by up-dissipated light that is reflected back to space (for example the dissipated light doesn't arrive at the world's surface).

A. Formation of cloud condensation nuclei by aerosol particles

Various kinds of airborne particles can go about as cloud buildup cores (CCN) around which the cloud drop-lets are shaped. The overall capacity of those distinctive kind of vaporized particles to supply cloud beads contrast according to their size and structure. The whole number and consequently such CCNs framed can straightforwardly impact the amount of precipitation, lifetime and along these lines the radiative properties of the cloud, and henceforth straightforwardly influence worldwide environmental change.



IV.EFFECTONHUMANHEALTH

Particulate issue is, at the present, being broadly explored because of its noticeable and undeniable effect on soundness of individuals, despite the fact that it really drew expanded consideration from researchers toward the beginning of the 80s. Numerous logical investigations have since then been distributed that shows that a prompt connection are frequently settled between openness to molecule contamination (both present moment and delayed openness) and well being chances, including unexpected passing. Such examinations demonstrate that fine particles are to a great extent a risk for the noticed medical issues in people and correspond openness to fine particles to expanded cases of asthma, diminished lung working capacity in kids, expanded clinic confirmations.

The hefty metal constituents of particulate is moreover a matter of significant thought. There are numerous significant confirmations that show that weighty metals adsorbed to environmental particulate is the thing that gives poisonousness to PM. These hefty metals have complex contamination highlights. Fine particles can contain weighty metals like arsenic, cadmium, selenium and so on. Numerous examinations on these metal constituents of PM have revealed that openness to high convergence of substantial metals can have a awful and dangerous effect on the human well being. moreover, these metals can get stored to plant leaves, soil, water bodies and so forth through wet and dry testimony. These may then get gathered inside the plant's body through biochemical cycles and people may collect these metals through utilization of the sullied plant.

Table 1 discusses some metallic chemical element that has a relatively high density and is toxic or poisonous at low concentrations and their associated effects on human health.

POLLUTANT	SOURCE	EFFECTONHUMANHEALTH	MINIMUMPERMISSIBLE LEVEL
ARSENIC	Chemical pesticides, electronic wastes, smelting etc.	Inflammation of the respiratory organs; bronchitis, dermatitis, inflammation of liver, anaemia, cardiovascular diseases	0.0003 mg/kg/day of oral exposure

LEAD	Emissions from vehicles and industries, paint, mining, burning of plastic etc.	Irritation of the gastro-intestinal tract, kidney and liver damage	Less than 10 micrograms per decilitre of blood
ZINC	Oil and petroleum refineries, plumbing, metal plating etc.	Irritation of the eye, suppression of iron and copper absorption	0.002-0.005 mg/kg/day
MERCURY	Electronic wastes, dental and pharmaceutical wastes etc.	Poor coordination, skin rashes, anxiety and memory problems, decreased intelligence	Less than 10 micrograms per decilitre of blood

Table 2: Effects of heavy metal constituents of PM on human health

POLLUTANT	SOURCE	EFFECT ON HUMAN HEALTH	MINIMUM PERMISSIBLE LEVEL
CADMIUM	Pesticides, chemical fertilisers, welding, nuclear power plant etc.	Gastrointestinal damage, kidney damage, has carcinogenic properties	Less than 1 microgram per decilitre of blood

III. EFFECT ON PLANTS

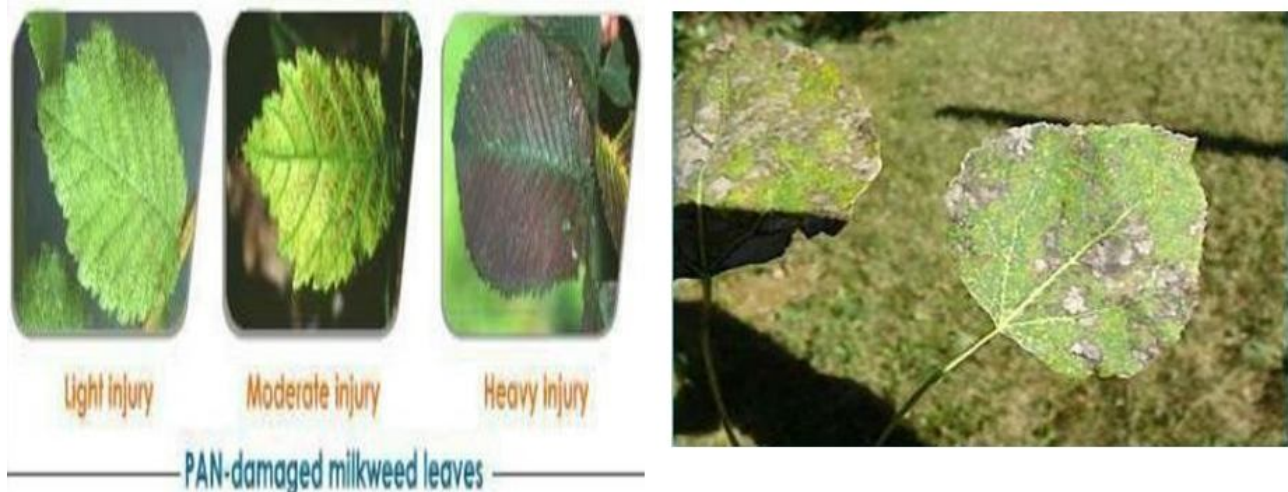
Different plants are delicate to air toxins and contaminations can harm their leaves, hinder plant development and breaking point essential efficiency. The most clear harm happens in the leaves. Development and propagation in certain plants might be disabled and the number of inhabitants in touch with species is diminished while open-minded species can flourish and overwhelm the vegetation. The significant harms brought about via air contaminations to plants incorporate chlorosis, corruption and epinasty. Plants may likewise decrease air quality, especially through the creation of dusts and spores which become airborne.

When plants are exposed to a number of pollutants, pollutants can cause different phytotoxic reactions in plants, taking into account the characteristics of the particulate mixture that is being deposited. Plants are typically capable of filtering coarse particles at a much higher rate than fine particles.

Particulate settlement on plants comprises nitrate, sulfate, trace elements and heavy metals. The acid-forming sulfate is relatively more soluble and the particle is more reactive than the mineral dusts. Particles with pH values ≥ 9 can prove to be

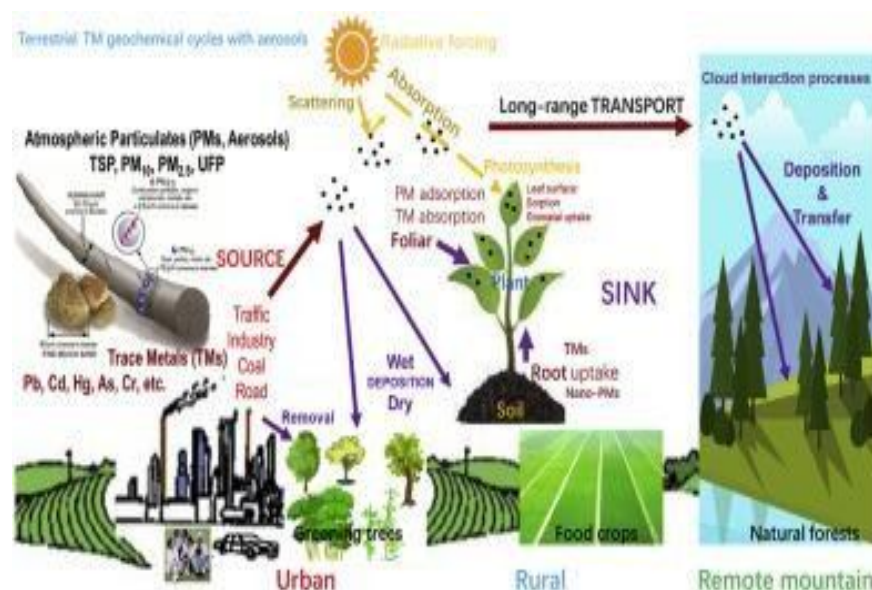
toxic to tissues of the leaf on which they are concentrated. They can also have a major adverse effect on plants indirectly by soil pH alteration.

Plants have a large leaf area for the absorption and deposition of dust particles. Dust accumulated on the leaves will dramatically change the optical properties of photosynthetic tissues and thus the amount of sunlight required for photosynthesis. This is when the dust particles are precisely arranged or target the surface of the leaf, essentially coloring or protecting the leaves from photosynthetically active radiation (PAR).



EFFECT ON ECOSYSTEM

The ecological reaction to particulate stress varies significantly. Succession in the unpolluted atmosphere is taken into account in a beneficial way, whereas certain areas where the pollution of particulate matter is high hinder the expansion of animals and plants. These particles interact with the conventional physiology and biochemistry of flora and fauna and consequently influence the organic phenomenon, the cycling of nutrients and the flow of energy. In addition, particulate matter from anthropogenic sources can harm the environment and organisms so badly that it becomes difficult for them to recover from the damage and propel them to further destruction. Chronic pollutant disruption to the forest environment may lead to a permanent depletion of susceptible species of plants and maybe even animals. Dust coating can induce abrasion and radiative heating, and can dramatically reduce photosynthetically active exposure to plant tissues.



CONCLUSION:

It is evident that particulate matter has major and consistent impacts on vegetation, the environment, human health and thus the ecosystem. Industrialization, growth and associated increase in energy consumption have resulted in a profound rise in these fine particles. The elevated amounts of these particles in the Earth's atmosphere will not only endanger the health of citizens, plants and animals, but also adversely affect the equilibrium of the environment. There is also every need to research, imagine and define potential techniques that can mitigate the presence of these particles in the atmosphere.

REFERENCES

- [1] Dr. Kedar M, Prof. SK Biswal, Effect of PM on plants, climate, ecosystem and human health, 1Department of chemistry GITA, BBSR, Odisha, 2IGIT, sarong, Dhenkanal, Odisha
- [2] LA Jimoda, Effects of PM on human health, the ecosystem, climate and materials; a review (2012), Department of chemical engineering, Akintolauniversity of technology
- [3] Christian Ul, Bettina W, Tanja MP, Arunava G, Inga M, Effect of solid particulate matter on vegetation (2008), 1Institute for horticulture sciences, Department of urban horticulture, 2Biological sciences division, Indian statistical institute
- [4] Santosh Kumar Prajapati, Ecological effects of airborne particulate matter on plants (2011), Department of botany, Guru GashidasVishwavidyalaya
- [5] Effect Of Particulate Matter (Pm) On Climate, Plants, Human Health And Ecosystems Rajpoot, K Lego International Journal Of Scientific Research 7 (5)
- [6] What is particulate matter? | <https://www.conserve-energy-future.com/causes-and-effects-of-particulate-matter.php>
- [7] Particulate matter | <https://www.citepa.org/en/air-and-climate/pollutants-and-ghg/particulate-matter>
- [8] Dr. Subhadra Rajpoot and DevangPratap Singh, “Emerging Public Health Concern and Air Pollution: A Case Study of Delhi’s Air Pollution Governance”, International Journal for Modern Trends in Science and Technology, Vol.06, Issue05, May2020, pp.:196-201