CHAPTER 4

REQUIREMENT ENGINEERING APPLYING FUZZY LOGIC

4.1 AIR QUALITY - SURFACE OZONE

The Fuzzy Logic approach is used in an air quality monitoring study at Chennai city done in A.C Tech., Anna University by Pulikesi (2005). This is a study of concentrations of various pollutants mainly surface ozone (O_3), and also others such as oxides of nitrogen (NOx), respirable suspended particulate matter (RSPM) and total suspended particulate matter (TSPM), under various climatic conditions viz. temperature, time of day, relative humidity, wind speed, wind direction etc. at different relevant geographical locations near the sea coast, low vegetation, far inland etc. with different usage patterns viz. industrial, residential, dense human population, high vehicle traffic etc.

Any representative pollution mapping and useful projection from this study to judge total air quality will need a complicated mathematical formulae involving many variables as above. This can of course be handled by a computer program processed in a lab PC. But such forecasting attempts involving climate factors have met with limited success since there are too many other variables which are relevant albeit in a much smaller way but still sometimes very effective. Further to establish the relationship to this pollution mapping to Public Health needs, they have to be related to the National Ambient Air Quality standards (NAAQS) which specifies the quality necessary with an adequate margin of safety for the protection of public health, sometimes for sensitive classes of the population, vegetation and property. By using Fuzzy Logic and it can be implemented in a simple process. This also proved that why this will be a more reliable process, since even after a large number of variables are accommodated in a mathematical formula, we can never have understood the interactions between these and also other unknown variables.

4.2 FAT THEOREM - FUZZY APPROXIMATION THEOREM WITH FUZZY ASSOCIATIVE MEMORY

Pulikesi (2006) investigated during the summer, concentrations of air pollutants. Different conditions of relative humidity (RH), wind speed (WS) and wind direction (WD) were collected over successive periods of about 24 hr at five sites, which were different in geographic characteristics and usage patterns. The readings as tables and graphical representation are given in Tables 4.1 to 4.3 and Figures 4.1 to 4.4.

Table 4.1 Details of air quality monitoring stations in Chennai, India

Site	Code	Latitude (N)	Longitude (E)	Site classification	Remarks
Kodungaiur	S_1	13.136851	80.2538	Industrial area	High Industrial emissions/ effluents
Koyambedu	S_2	13.07224	80.20174	Commercial area	Service Industries/ Interior
Mandaveli	S_3	13.02596	80.26618	Commercial area	Service Industries/ Coastal
Taramani	S_4	12.98112	80.23949	Residential area	Dense population
Vallalar Nagar	S_5	13.10571	80.28016	Traffic island	High motor vehicle emission

The National Ambient Air Quality standards (NAAQS) has specified quality necessary with an adequate margin of safety for the protection of the public health, vegetation and property. The NAAQS Standards for the air pollutants i.e. nitrogen oxides (NO_x), Respirrable Suspended Particulate Matter (RSPM) and Total Suspended Particular Matter (TSPM) are summarized in Table 4.2.

Table 4.2 National ambient air quality standards (NAAQS) air pollutants

Pollutants	Averaging Time	Concentration in Ambient Air (ppb)		
		Industrial Area	Residential, Rural and other Areas	Sensitive Area
Oxides of Nitrogen NO_2	Annual Mean	40.8	30.6	7.65
	24 Hours	61.2	40.8	15.3
Respirrable Suspended Particulate Mater (RSPM)	Annual Mean	183.6	71.4	35.7
	24 Hours	255	102	51
Total Suspended Particulate Matter (TSPM)	Annual Mean	61.2	30.6	25.5
	24 Hours	76.5	51	38.25

Table 4.3 Pollutant readings in different areas

Kodungaiur (S_1)		Pollutant (ppb)		
Time	O_3	NO_x	RSPM	TSPM
06-14	37	18.3	35.19	139.74
14-22	22	9.02	28.56	52.02
22-06	18	3.21	18.87	33.15
24 Hrs. Avg.	25	10.2	27.54	74.97
Koyambedu (S_2)		Pollutant (ppb)		
Time	O_3	NO_x	RSPM	TSPM
06-14	28	12.03	20.4	123.42
14-22	21	9.84	31.62	117.81
22-06	8	9.33	11.73	28.05
24 Hrs. Avg.	18	10.4	21.42	89.76
Mandaveli (S_3)		Pollutant (ppb)		
Time	O_3	NO_x	RSPM	TSPM
06-14	38	12.64	20.4	107.61
14-22	35	10.25	17.34	59.93
22-06	22	9.58	24.48	61.2
24 Hrs. Avg.	31	10.81	20.91	76.5
Taramani (S_4)		Pollutant (ppb)		
Time	O_3	NO_x	RSPM	TSPM
06-14	36	4.69	29.58	110.67
14-22	40	4.43	28.56	90.27
22-06	18	2.6	24.99	55.59
24 Hrs. Avg.	30	3.92	27.54	85.17
Vallalar Nagar S_5		Pollutant (ppb)		
Time	O_3	NO_x	RSPM	TSPM
06-14	10	27	87.72	212.16
14-22	16	15.81	24.99	117.81
22-06	4	8.16	33.15	76.5
24 Hrs. Avg.	10	16.93	48.45	135.66

In Figures 4.1 and 4.2 hourly O_3 concentrations are shown as a function of relative humidity. The other meteorological data i.e. wind speed and wind direction was available later on that day. In the early morning hours, O_3 concentration showed a sharp increase in ambient air and reached its first peak in 11.00-14.00 as also recorded by National Research Council (NRC). The first peak occurred at temperature of 42^{o}C and relative humidity of 47% (Figure 4.1). Subsequently decrease of O_3 concentration may be due to the cloudy weather conditions. As mentioned by NRC a second peak was also observed in the 16.00-20.00. After that, a gradual decrease of O_3 concentration continued until the early morning of the next day.

In Figure 4.3, hourly variations of O_3 concentrations are shown as a function of meteorological parameters obtained i.e. (a) Ozone and wind speed, (b) Ozone and wind direction, (c) Ozone and relative humidity.

In Figure 4.4, hourly variation of ozone concentrations are shown as a function of meteorological parameters obtained i.e. (a) Ozone and wind speed, (b) ozone and wind direction, (c) Ozone and relative humidity.

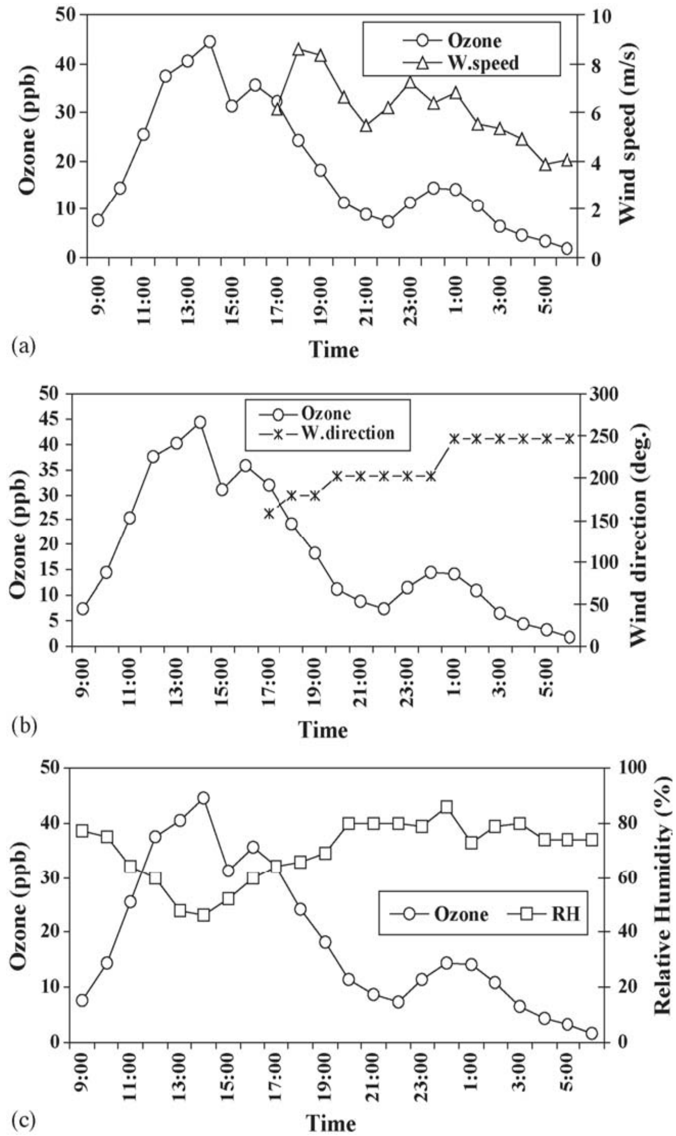


Figure 4.1 **Hourly variations of ozone and meteorological parameters obtained from 25 to 26 May 2005 at Koyambedu. (a) Ozone and wind speed, (b) ozone and wind direction and (c) ozone and relative humidity**

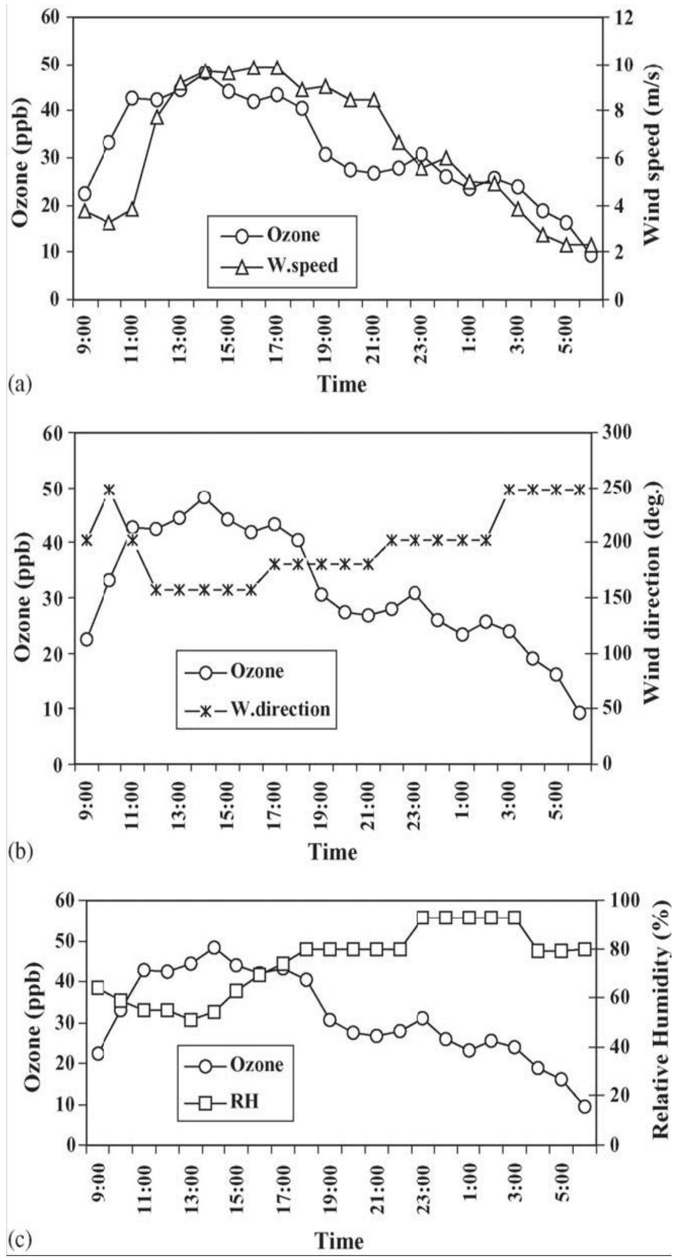


Figure 4.2 Hourly variations of ozone and meteorological parameters obtained from 31st May to 1st June 2005 at Mandaveli. (a) Ozone and wind speed, (b) ozone and wind direction and (c) ozone and relative humidity

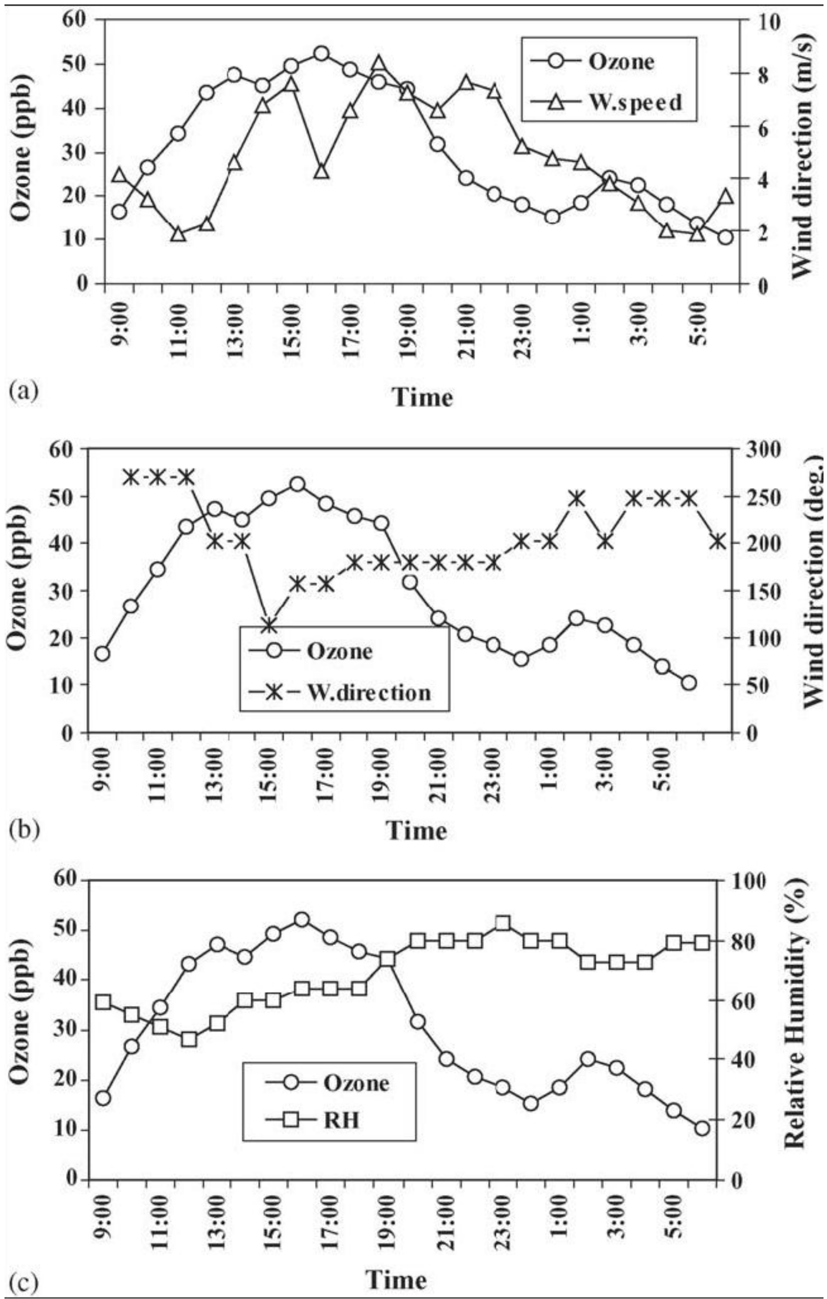


Figure 4.3 Hourly variations of ozone and meteorological parameters obtained from 03 to 04 June 2005 at Taramani. (a) Ozone and wind speed, (b) ozone and wind direction and (c) ozone and relative humidity

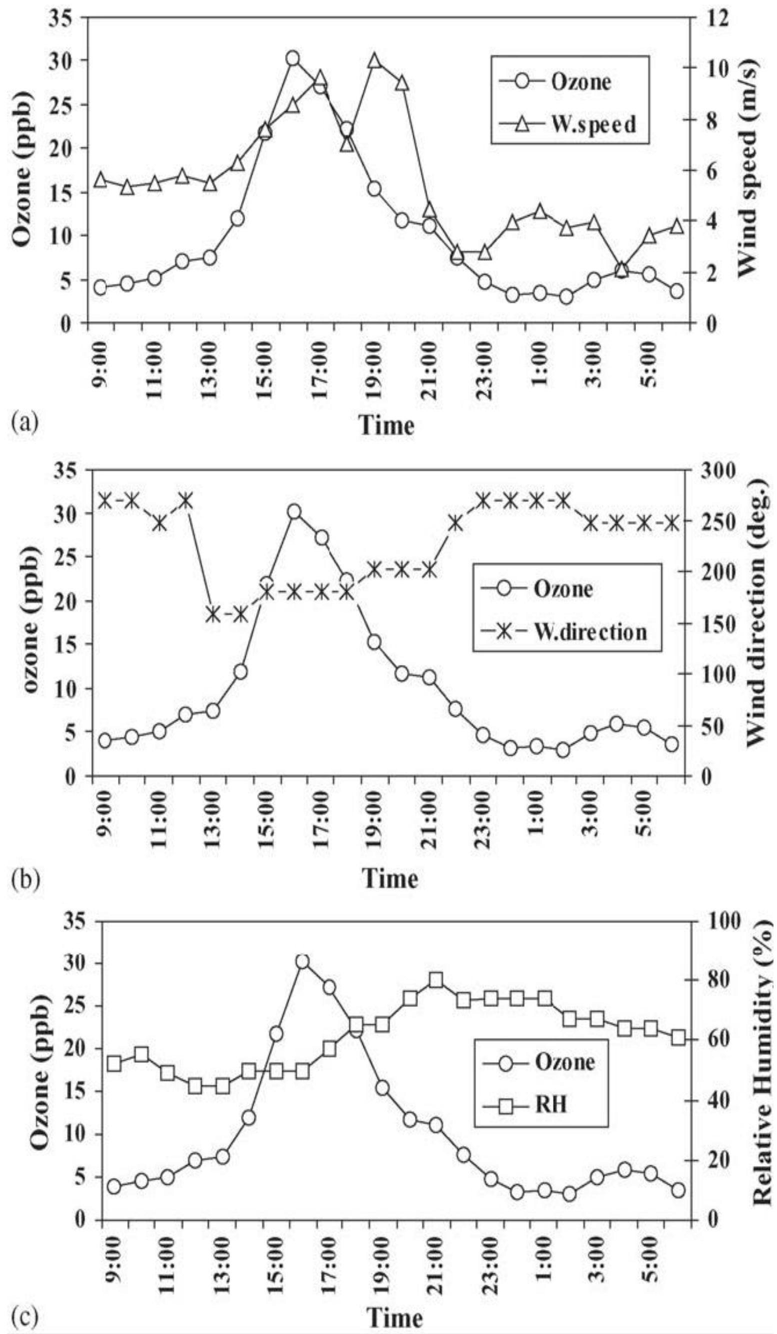


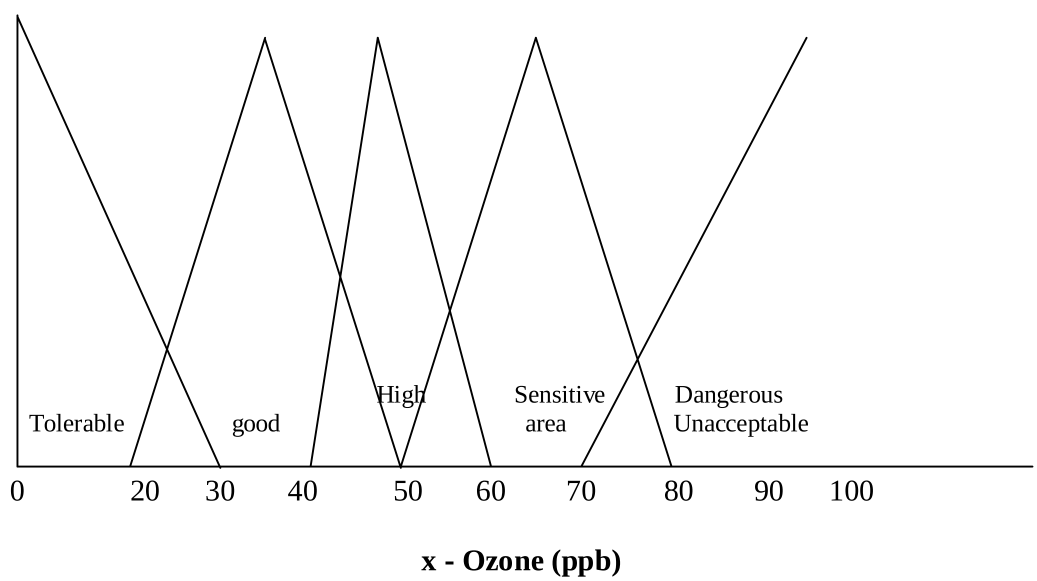
Figure 4.4 Hourly variations of ozone and meteorological parameters obtained from 04 to 05 July 2005 at Vallalar Nagar. (a) Ozone and wind speed, (b) ozone and wind direction and (c) ozone and relative humidity

In a digital system, one can relate the changes in concentrations of each pollutant to each condition separately or one can derive mathematical formulae to try and relate the pollutant concentrations to changes in all these conditions at each time. These formulae will be very complicated and for easier handling one may have to approximate to lower degrees. When working in the reactions between each of these pollutants at different concentrations, the formulae become even more complex. In the fuzzy system, we first establish rules or trends in the variation of each of these pollutants either from experiences of experts in the field or by an analysis of readings or also other similar work.

4.3 POLLUTION MAPPING AND FORECASTING MODEL

The need is a program to give results based on which decisions can be made for Locating population centers, industrial areas for polluting or non-polluting processes, outdoor entertainment sports sites etc. The result can also be on a weighted average of the acceptable limits allowed by the National Ambient Air Quality standards (NAAQS) of all the pollutants studied. The inputs into the program would be the various readings taken as done in the study. More the readings taken over longer periods would give more reliable results. A process known as Fuzzy Associative Memory, which is explained below in detail, merges the data for each pollutant under different conditions. The resultant data for each pollutant is again fed into another module to compare with the allowable standards of NAAQS where we use another FAM to arrive at an indication of Total Air Quality.

In this analysis example at first we shall take only one pollutant ie Ozone (O_3) is taken and track its variations against only one of the climatic variables i.e. relative Humidity (RH). Now pick the variables say x – O_3 and y – RH. A geometric representation is used rather than mathematical symbols as this more easily striking to a larger number of people in the climate forecasting and pollution study fields. Figure 4.5 below is a representation of the foll statements WHO standard for permissible high O_3 concentration is 50 ppb i.e. from 40 to 60 can be considered the high region with 50 ppb being definitely only in the high region and 40ppb in the high as well as in the good region. Similarly, the other areas signify their membership in a scale in similar representation which captures the real fact that each reading can be "somewhat" in other regions also.



WHO permissible - 50ppb

Figure 4.5 Ozone distribution

Similarly, Figure 4.6 depicts the Relative Humidity regions of comfort, humid, dry etc. with the NAAQS being 60% for maximum comfort.

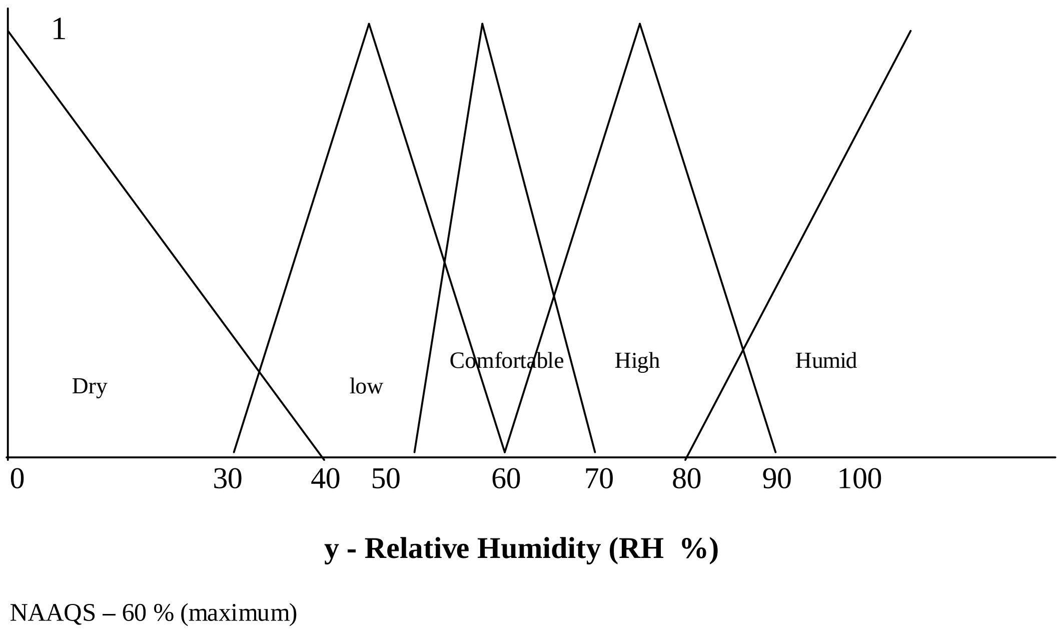


Figure 4.6 Relative humidity distributions

Figure 4.7 depicts the region where the Rule "if RH is high then O_3 is good" is satisfied. After picking the variables x and y, pick the fuzzy sets for the variables x and y, i.e. here for O_3 and RH. The fuzzy sets chosen for O_3 are tolerable, good, high sensitive area and unacceptable / dangerous. The fuzzy sets chosen for RH are dry, low, comfortable, high and humid after choosing the appropriate fuzzy sets to find the rules. These are found from trends in readings expert opinions and also adaptive learning systems where available. On study of the readings taken and discussions, one can assume many trends. Some are given below for example.

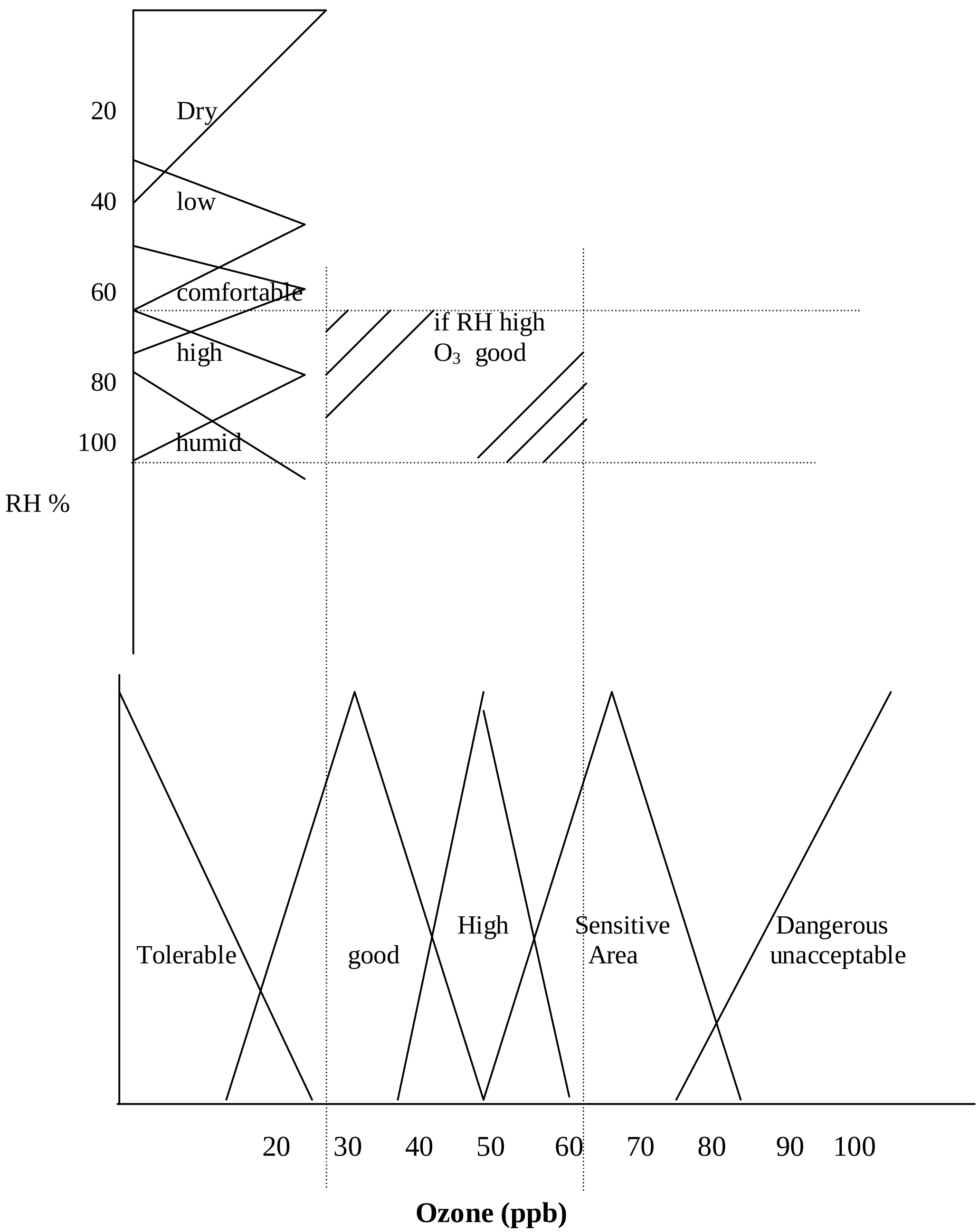


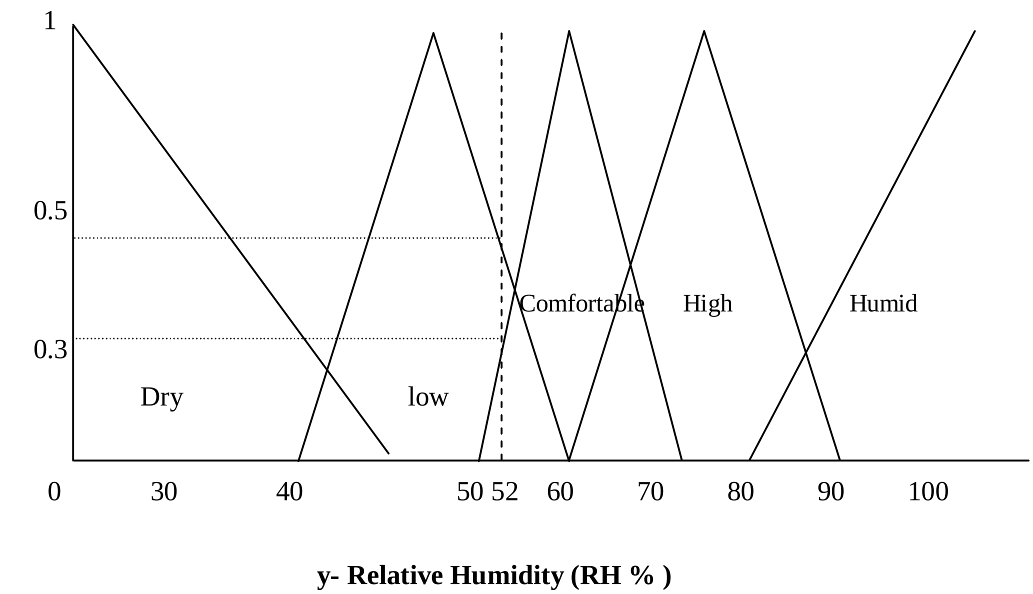
Figure 4.7 Ozone to relative humidity

1. When RH is high O_3 is low- mostly inversely proportional provided all other factors are same.
2. When temperature is high NOx emissions from soil are high.
3. Wind direction from the sea brings in O_3 and hence concentration increases.

These trends are also obvious from a casual study of the graphical representation of the comparative readings in Figures 4.1 to 4.4. From such trends and discussions with domain experts we formulate rules. As an example following five rules can be assumed.

1. If RH is dry then O_3 is dangerously high and unacceptable.
2. If RH is low then O_3 is higher than max and unacceptable to sensitive sections like aged and sick.
3. If RH is comfortable then O_3 is high around WHO max.
4. If RH is high then O_3 is in the good region.
5. If RH is humid then O_3 is low and not healthy, but tolerable.

A graphical representation of the above rules can be seen in Figure 4.7 above. Rule 4, if RH is high then O_3 is in the good region applies to all the points lying within the bounded shaded area. Such rules can be formulated independently for pairs of all other variables, which have a bearing on O_3 concentration such as Wind direction to O_3, NOx to O_3, temperature to O_3, seasons – winter, summer etc. to O_3 etc. with similar geometric representation. From Figure 4.8, we see that a reading of RH say 52 can lie in both the low and comfortable regions each to some extent. Hence, for a RH reading of 52 both Rule 2 and Rule 3 will fire simultaneously each to a certain extent. One can accommodate this in fuzzy logic but not in digital logic where it is EITHER – OR. To get a representative possible value of $O_{3,}$ we use the FAM. This is explained below and represented geometrically in Figures 4.8 and 4.9.



NAAQS – 60 %

Figure 4.8 RH 30% comfortable - 50% low

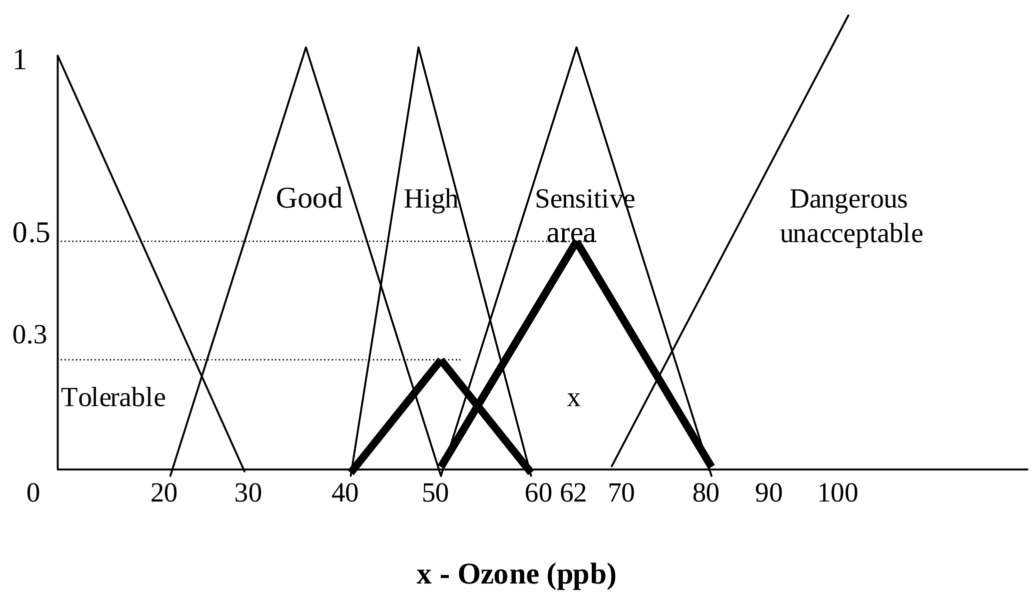


Figure 4.9 "x" Geometric centroid

From Figure 4.8, one can notice that when RH is 52 it is 30 % in the comfortable region and 50 % in the low region; i.e. Rule 2 kicks in for 50% effect and Rule 3for 30% effect. This is shown in Figure 4.9 where the effective weight in the corresponding regions of O_3 are shown, i.e. there is a 30% effect in the high O_3 region and a 50% effect in the sensitive area. This can also be called the Degree of Belief in a conclusion. Weighted Inference Equations (basic Neural Equations) can be used to obtain the defuzzified result. However, the geometric representation, explained in detail in chapter 3 (The Kosko method) is more intuitive and hence used here. In the geometric representation this is achieved by shrinking down the corresponding triangles by the respective values as shown in Figure 4.9.

The value of O_3 is the resultant fuzzy set under the dark line triangles i.e. O_3 value can fall anywhere in the area. However to get the most probable representative value one need to defuzzify this result. It has been found that taking the centroid of the area under the dark line gives the best-defuzzified value. The centroid at point “x” is at the value 62 ppb. This is the defuzzification method found to be the most relevant and accurate in practice. We need this crisp result to feed into a computer for further processing. Similarly, one can take a fuzzy weighted average for each of the climatic conditions, which affect the ozone concentration. All these weighted averages can again be fed into a FAM module as shown in Figure 4.10, to get the final defuzzified figure which will be the most likely O_3 concentration for the values of variables fed in.

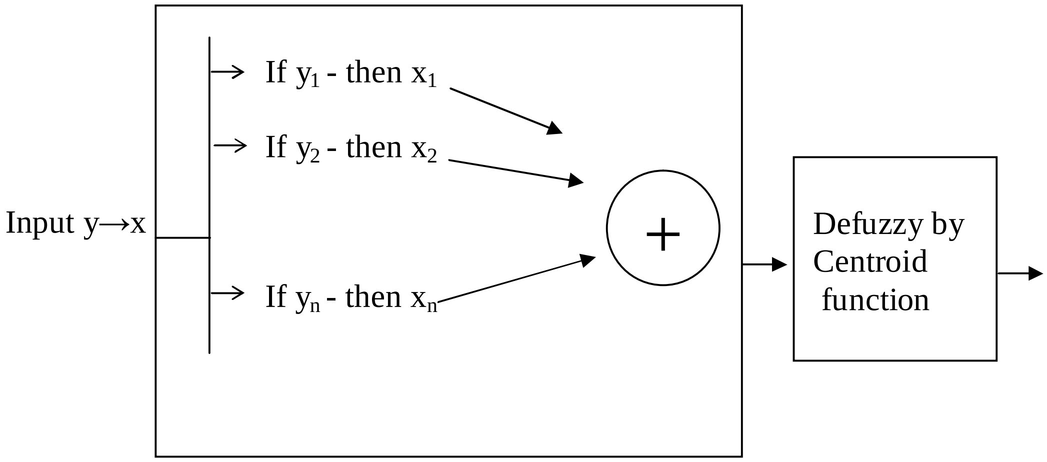


Figure 4.10 Fuzzy associative memory

4.4 ADAPTIVE SYSTEMS

The FAT theorem (Kosko 1993) tells that in theory one can always find fuzzy rules to simulate or approximate any type of control or computer processing. But in practice, one may have no idea where to begin. The FAT is more helpful. It suggests ways to automate the search for good rules. Neural networks or brain like computer systems can help find these fuzzy rules. A system that learns fuzzy rules from experience. That means it learns from data. Fuzzy engineers turn expert talk into fuzzy rules. They use the judgment that hides in their brains. We seek a learning system that turns expert behavior into fuzzy rules. The experts leave a trail that the adaptive system converts into fuzzy rules. The rules leave a similar trail in the data. The adaptive system sucks the brain of the expert or the computer or whatever came up with the data. The more data, the better the brain suck, the finer the fuzzy rules. Data to rules.

DIRO: Data in, rules out. Numbers in, knowledge out. Experience in, common sense out. Examples in, expertise out. A neural network can fill the black box (Figure 4.11).

Figure 4.11 DIRO box

A neural net acts like the eyes and ears of an adaptive fuzzy system, a fuzzy system whose rules change with experience. The neural net senses fuzzy patterns in the data and learns to associate the patterns. The associations are rules: If fuzzy set A, then fuzzy set B. The Fuzzy system acts at a higher cognitive level to reason with the fuzzy rules. It infers or decides and outcome action based on the incoming data or facts. An adaptive or neural fuzzy system changes or tunes its rules as it samples new data. Just as every pattern we see or hear or taste or feel changes slightly the world view, so every new example of expert behavior changes slightly the rules in an adaptive fuzzy system to find a rough working set of fuzzy rules. Then with more samples, more expert examples, the rules change more slowly as the fuzzy system fine-tunes its knowledge. With neural nets and with brains practice makes perfect.

One does this until it is right without thinking about it until it feels right. Practice makes perfect by making a habit and then refining the habit. Skill improves a lot at first and then improves slowly and then does not improve at all. It grows and then “max” out. First a run up the “learning curve” and then only a crawl up. One needs no verbal skill to learn to swim. Just have to practice it. Most natural swimmers cannot explain how they learnt but they are glad to show.

Take a set of data in a graphical representation. One can find clusters where data bunches up. The clusters in the data are like patches Learn data clusters and one learns patches and one learns rules and that is an adaptive fuzzy system. That's the answer but it depends on learning. This learning depends on neural nets.

To learn is to change. And to change is to learn. You can learn well or badly. But you cannot learn without changing or change without learning. Your brain changes. It changes a little bit every time you see an image or hear a sound or feel a surface or taste a flavor or walk a new ground. Everything you sense changes your brain. Your brain measures things and those things change your brain. It learns new changes and forgets or unlearns old changes. The neurons in the brain do not act as computer memory sites. No cell holds a picture of your house or the smell of lime or the idea of God. You can pull out any cell in your brain and your mind will not change. You could pull out a few million cells at random and not miss them. Pull a few wires or circuits out of a computer and it crashes.

What counts is the pattern, the synapses or neural connections. Each neuron in your brain can connect up to 10,000 other neurons. Learning and memory lie in the great tangled webs of synapses. Not in cells but in webs of neurons and synapses Learning is change. In brains that mean learning is change in a synapse. So learning changes synapses. That still does not tell how to encode or decode a pattern in a neural net. Each net has thousands or millions of neurons in it. And each neuron does nothing but connect up with other neurons through synapses by chemical/electrical squirts. A pattern is in whole fields of neurons. Language is made up of sentences made up of words made up of syllables or phonemes. In brain language, the syllable is a pattern of activation, a whole field or slab of neurons that reverberate or resonate. The power of neural nets is that they self-organize. And may be our brains act the same way. In 1977, Shun-ichi Amari of Tokyo University found the math that describes how nets of simple on-off neurons behave.

Stephen Grossberg of Boston University has shown that we learn a new idea or pattern only if it resonates with what we expect to see or hear or think. He has extended “adaptive resonance” theory to the learning of fuzzy-set concepts. Synapses work the memory or lose them and the more you use them the bigger they get. The memory changes as the synapses change.

The big point is that learning changes an information medium. In our brain, it is a web of synapses or webs of synapses or webs of webs of synapses and maybe some muscle-contraction rates. The information medium can be anything. A computer learns when you change its software or memory circuits. Warm wax learns your palm print when you push your hand into it. The canvas learns the painting when the artist smears paint on it. Even your lawn of green grass can learn what you mow it. This odd case gives a good example of learning in a parallel information medium. So rig or “train” with samples a neural net to find the clusters. That means data clusters. The more examples the better. Data in, rules out. The example shows a second and related way to use neural nets to grow rules from clusters.

Fuzzy rules are patches. The next step ties patches to data clusters. This gives new force to the FAT theorem that says a fuzzy system can model any system by covering its system curve with small fuzzy patches. With enough data, the fuzzy system can learn any system. Adaptive fuzzy systems work at two levels. Each level approximates something. At the small or local level a neural net approximates patches or rules. At the big or global level the patches approximate the whole system. Points make up patches. A patch will grow more points in a region, if we make it denser, if we form a cluster of points. More points help the neural net.

In a few hours of data, the square is filled with a swarm of points. The swarm is a data cluster. It shows how an expert answers questions or associates outputs with inputs. As David Hume the Scottish Philosopher saw, it shows how we associate similar outputs with similar inputs. It defines a rule. This learning scheme is product space clustering. It just means that a rule is a cluster in the data. More clusters, more rules. Each neuron has its own web of synapses that flow into it. The neurons “complete” a web as each new piece of data rolls in. The new data also define a point in the geometry. Then it gets to learn by changing its synaptic web a little so that it looks more like the data.

You quantize when you round off or when you pick an example. Quantization compresses information. The data points grow and grow in number. Learning never stops. Rules are reinforced where the data points are dense and sparse where the data points are sparse. This means the new dots estimate data clusters. So they estimate fuzzy rules, since rules are patches and clusters cover patches.

You learn rules with such a net. Just count the big black dot clusters in the rule cells. If a cell has a Dot Cluster in it, add that rule to the system. You can weight the rules this way too. Some cells have more Dot Cluster than other cells have. This can mean they are more important to the expert or it can just mean that the expert hangs out in that region of the plane. In practice, at least two Dot Cluster per cell is needed to count as a rule. This cuts down on bad rules that can come from noisy data or from a rare bad example. Even the best experts make mistakes.

A Dot Cluster net fills in the DIRO black box: Data in, rules out. It slowly sucks the expert’s brain as one feeds it expert data. At first the Dot Clusters do not spread out well. That gives a small number of bad rules. Then with more data the Dot Clusters spread out to track the data. That gives more and better rules. More data refine the spread of Dot Clusters and polish the rules. At this point one can drop the human expert and just work with the fuzzy rules. They will give the same behavior. In practice, we then polish these rule by playing with the size and position of the fuzzy sets or by letting a supervised neural net tune them. Then the tuned fuzzy system can outdo the expert.

There are many ways to grow fuzzy systems from data. They all boil down to clustering data into rules. So for this problem, the focus is on that idea. A few dozen other approaches that in the end do the same thing is also possible.

These make up the new field of adaptive fuzzy systems, or neuro-fuzzy systems, or fuzzy neural systems. The neural nets just use data to find or tune rules. The rules are patches and they cover a wiggling system curve in accord with the FAT theorem.

4.5 AIR QUALITY - SURFACE OZONE

We use the data from the Ozone study by M.Pulikesi et.al to validate this technique. The problem here is to forecast environmental characteristics like Relative Density (RH), Ozone levels, RSPM etc. with respect to other given parameters like time of day, temperature, pollution levels of Nox etc. Because of the large number of parameters affecting any of the characteristics and the interaction between the parameters we can never have exact forecasts. However, larger and larger numbers of readings can give more reliable forecasts.

4.6 ADAPTIVE FUZZY SYSTEM FOR PREDICTING ENVIRONMENTAL CHARACTERISTICS

This is an attempt to use an adaptive fuzzy system, which automates rule discovery and continuously refines the system to improve upon the result. In the studies done by Pulikesi during 2005, readings of various environmental / pollution characteristics at different areas of Chennai city were taken as shown in Table 4.3. As an illustration of the method, a study at first of only the relationship between two parameters was taken up for each place and then the study to integrate other parameters also. The readings of Ozone against time in the five areas under study was taken first. Next the readings of Ozone against RH are taken. The graphs below show some of the readings. Even though a very exhaustive study was done only a few for illustration are taken up.

The dots in the Figure 4.12 are the various readings taken on different days but at the same time i.e. 10 am 12 noon, 2 pm, 4 pm etc. on many days. The readings form a kind of fuzzy cloud around a region of the graph. These clusters of dots gives an understanding of what ozone levels can be at various times of the day to a human observer i.e. from the DOT clusters seen in Figure 4.12 the foll Rules can be inferred

1. Ozone peaks in the afternoons between 12.00 noon and 16.00 hours.
2. Nights are the troughs from 20.00 to 8.00 am.

As the number of observation go on the dots get clustered thicker in some areas and sparser in some areas. This reinforces or weakens the rule. To work on these for further processing in a FAM etc. this has to be fed in a digital computer .For this we need a crisp value i.e. we quantize or approximate the dots. A plain center of gravity of a cloud gives a representative crisp value.

Figure 4.13 is the rounding off/quantizing of the DOT clusters seen in Figure 4.12.

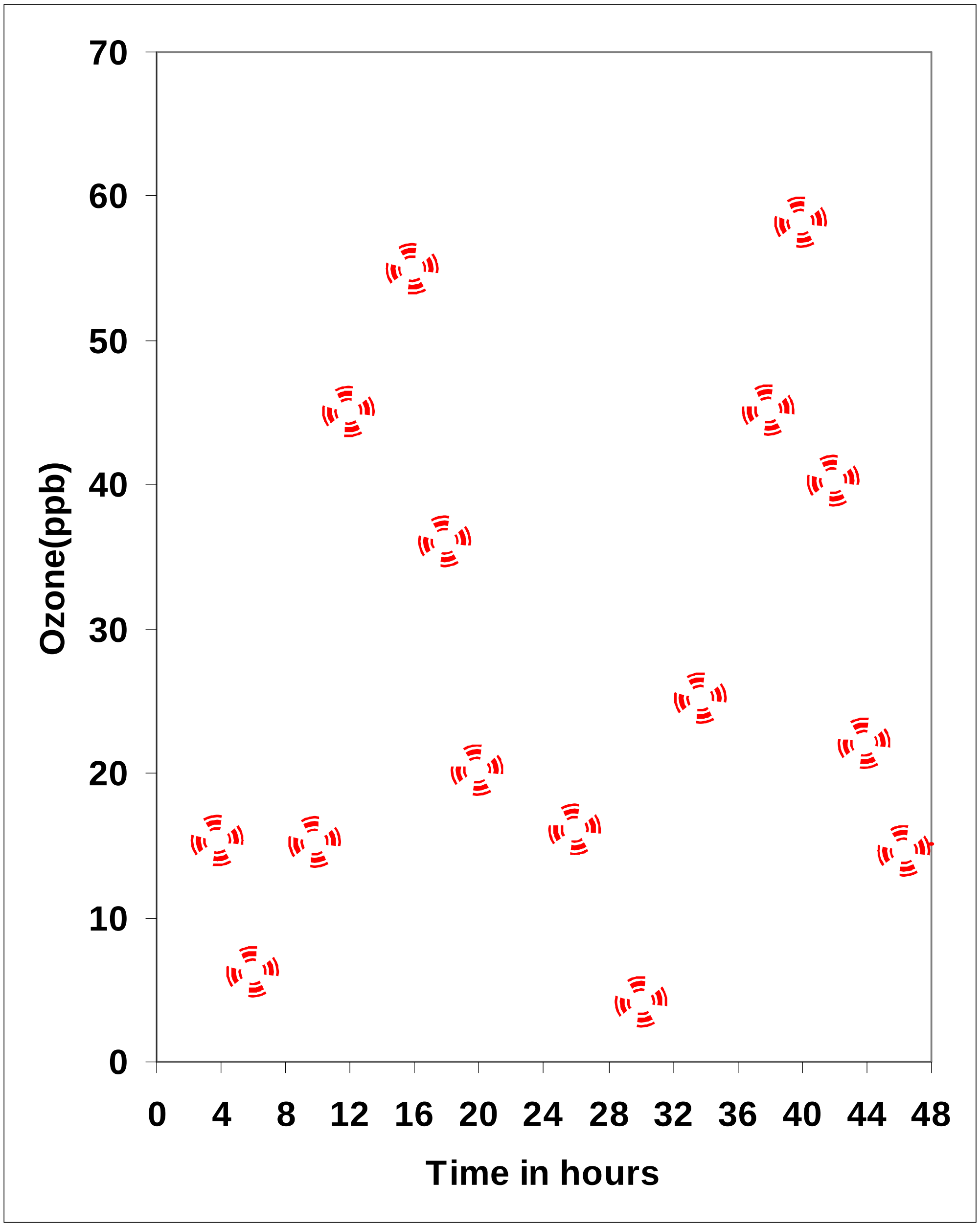


Figure 4.12 Ozone readings on many different days at a particular place on hourly basis – Dot clusters

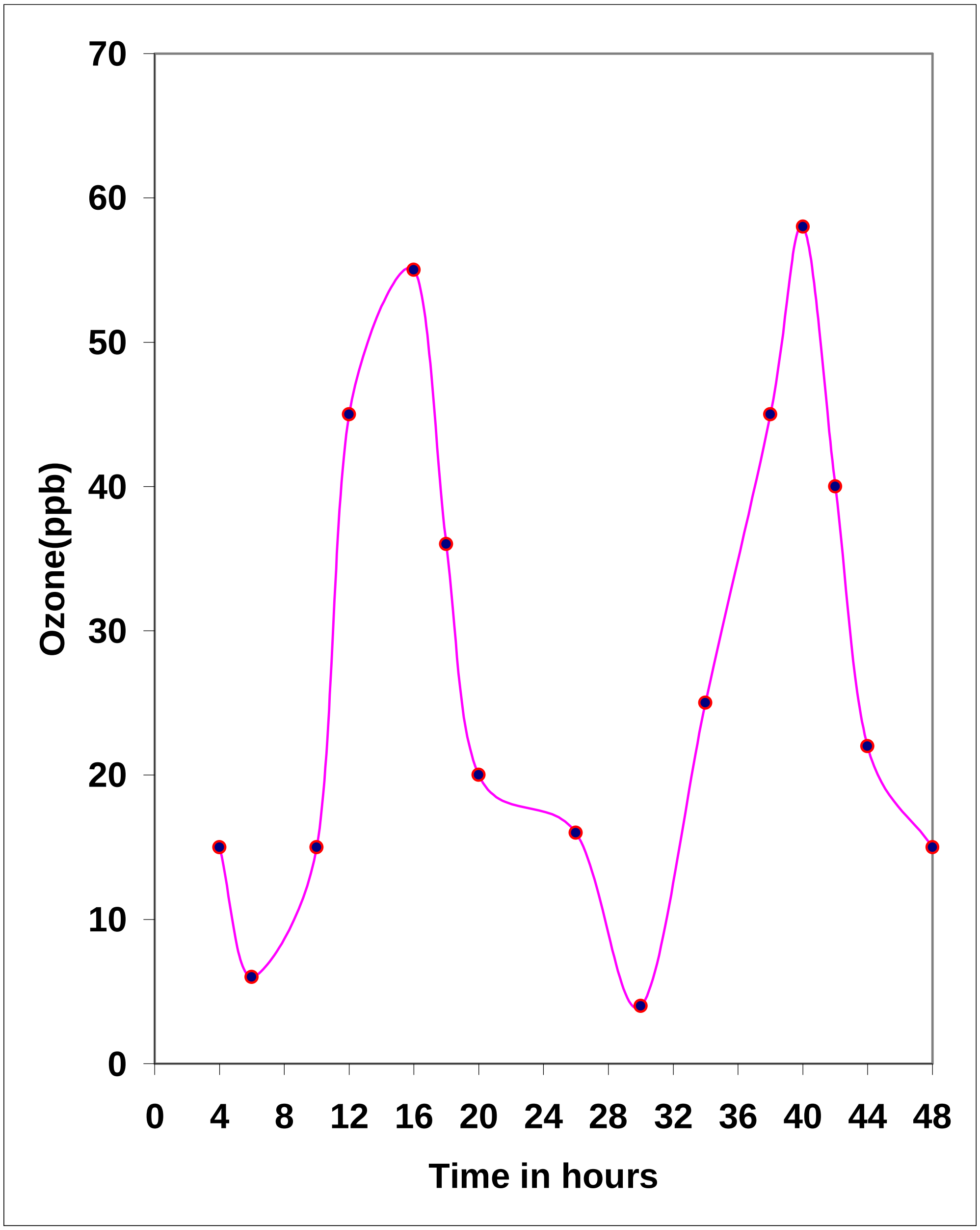


Figure 4.13 Quantized of Dot clusters from Figure 4.12 and graph showing trend

The Dots in Figure 4.14 are the various readings of ozone at same RH of different times but in same place. This is taken from many ozone and RH readings taken at the same place and grouped suitably to relate only these two parameters. From the DOT clusters seen in Figure 4.14 the foll rules can be inferred.

1. If RH is dry then O_3 is high.
2. If RH is low then O_3 is higher than WHO max
3. If RH is comfortable then O_3 is high around WHO max.
4. If RH is high then O_3 is in the good region.
5. If RH is humid then O_3 is low and not healthy, but tolerable

The linguistic terms – RH is dry, low, comfortable and O_3 is high, higher than WHO max. etc. represent the regions as explained earlier in Figures 4.5 and 4.6. Similarly, Figure 4.15 is the rounding of the DOT clusters seen in Figure 4.14 and graph showing the trend. As more readings are available during future studies more dots are generated which are rounded off to produce closer points for a smooth graph. Thus we write a programme, which generates more refined rules by itself. To incorporate more climate parameters like NOx, wind direction etc. we can use a FAM as explained earlier and represented in Figure 4.10. A self-refining programme incorporating new or stronger rules and dropping weaker rules is possible with this method. Since the relationship between parameters (which are the rules) is validated of each time with the actual resultant weather condition this is a continuously improving system.

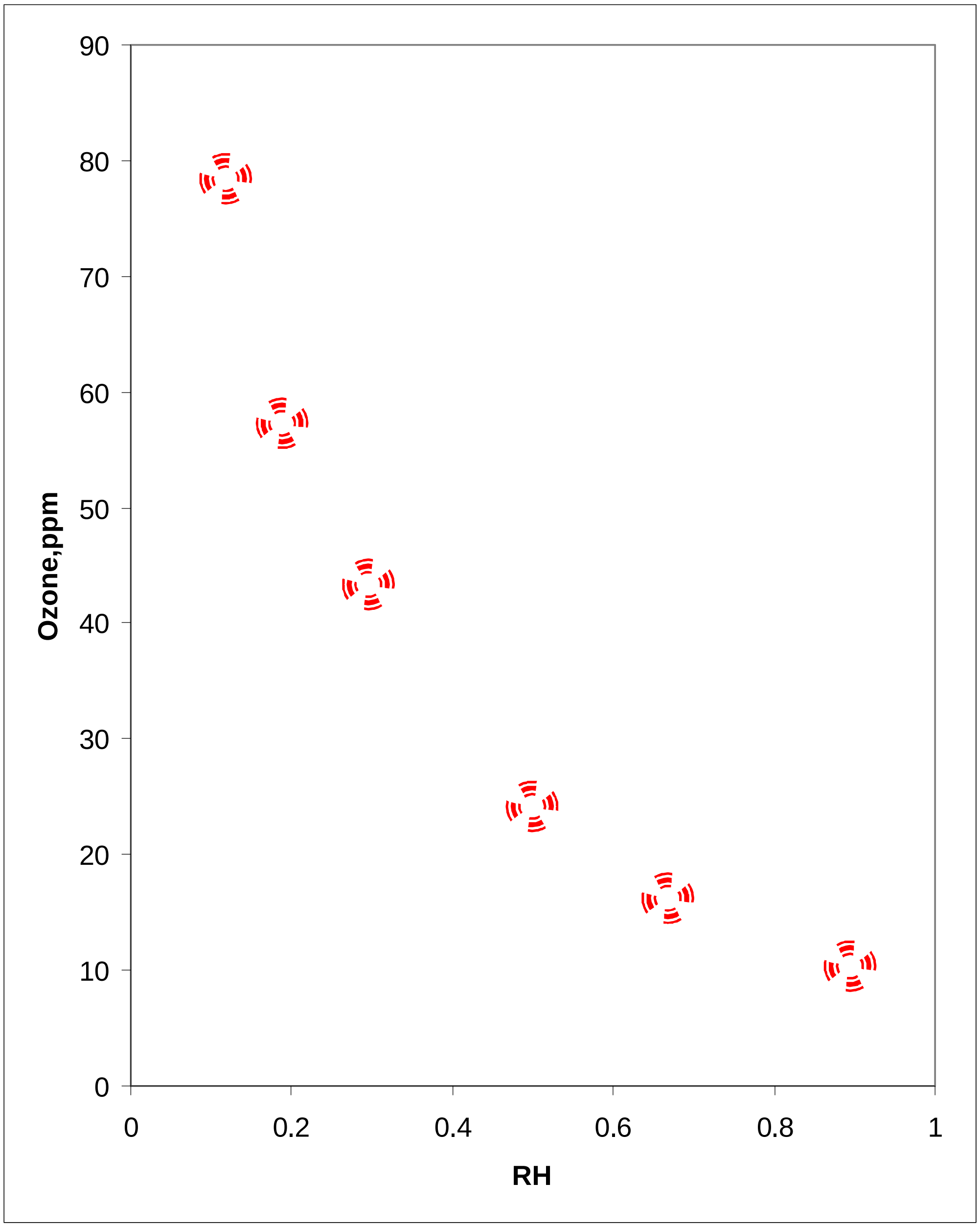


Figure 4.14 Ozone readings against relative humidity on different days at same place – Dot clusters

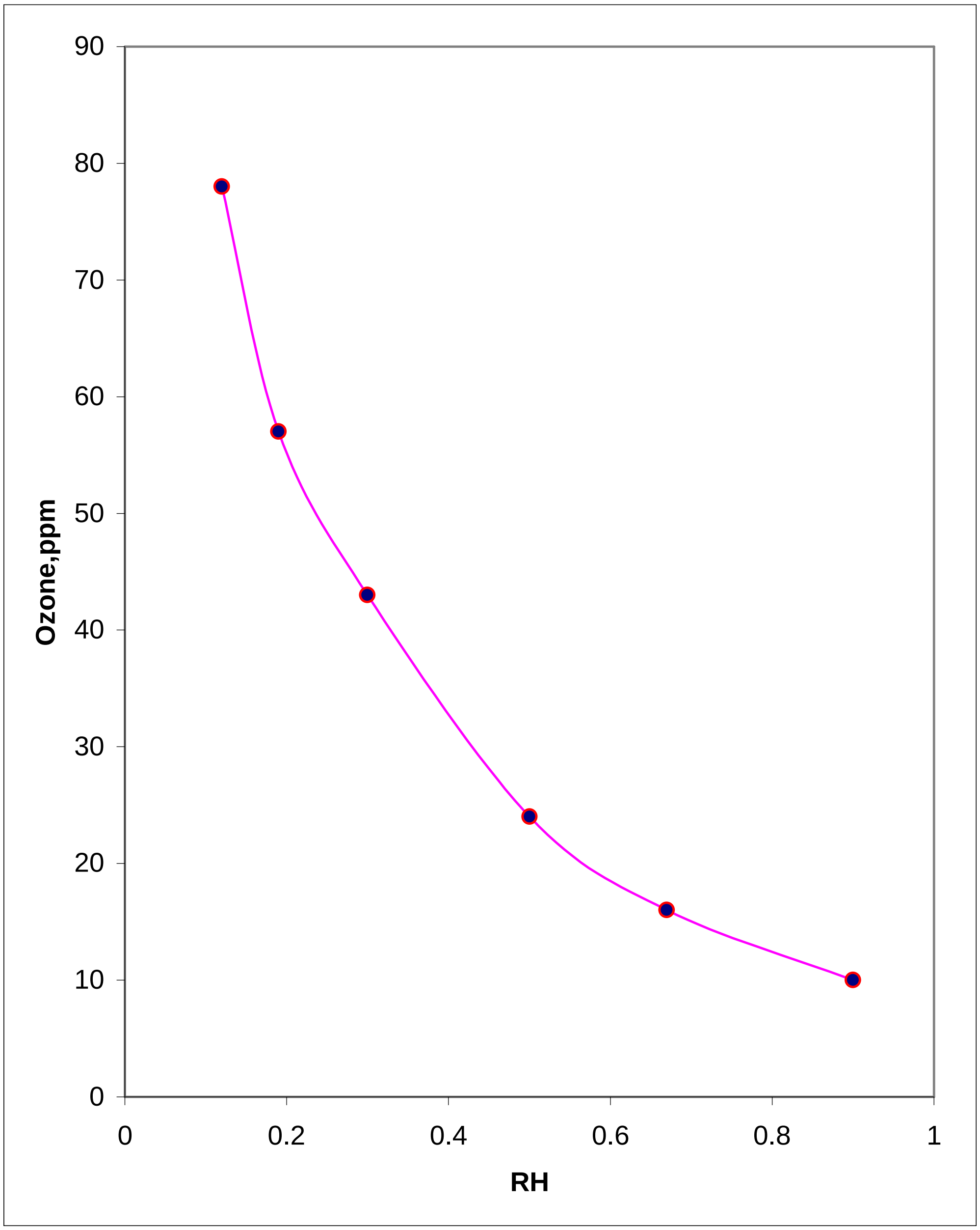


Figure 4.15 Quantized Dot clusters from Figure 4.14 and graph showing trend