Creating Human-Based Sensory Networks through Crowdsourcing Techniques in Transportation Studies

Francis Aldrine A. UY [a], Jocelyn Buluran[b], Farkam MOHEBBI [c],

[a,b, c] *School of Civil, Environmental and Geological Engineering, Mapua Institute of Technology, Muralla Street, Intramuros, Manila, Philippines 1002*
[a] *E-mail: faauy@mapua.edu.ph*
[b] *E-mail: jsbuluran@mapua.edu.ph*
[c] *E-mail: fmohebbi@mymail.mapua.edu.ph*

Abstract: Crowdsourcing online is one of the latest techniques of gathering data or opinions in surveying from a large group of people. This technique is cost-effective in developing countries like the Philippines that have no sufficient funding for development projects. The use of this technique when found to be reliable would significantly lessen cost and labour in data gathering. This paper presents the potential of crowd sourcing technique in creating human-based sensory networks for transportation studies. A number of cases that are based on initial studies are presented in this paper. The crowd sourced data were compared to actual observations or records. The results of initial studies confirm the potential of this technique in conducting cost-effective transportation studies.

Keywords: Crowdsourcing, Survey, Online Community

1. CROWDSOURCING

Crowdsourcing is the practice of gathering data or ideas by surveying from a large group of people, and especially from an online community. This process can occur both online and offline, but mostly use online because of ease of time and the cost effectiveness. The general concept is to combine efforts of a large number of volunteers in a community where each one can contribute a small portion and at the end significant result can be found from the data.

The most common crowdsourcing activity that the Philippines encounter is from the government where their employees goes to houses and asks a few questions about how many people living there and so on. Lately, this crowdsourcing from the government have sprouted and started to be used by different agencies like Metro Manila Development Authority (MMDA).

Crowdsourcing would greatly help to gather information on a large scale quickly and the information obtained are fresh and reliable, and beside of those advantages, crowdsourcing promotes values like cooperation, honestly and unity among citizen along the areas of active crowdsourcing.

2. CROWDSOURCING TECHNOLOGIES

Currently, many types of crowdsourcing technologies are available to facilitate this issue, but the most common technology which used for crowdsourcing are computers and electronic devices like tablets, mobile phones and any device which can connect and transfer the data to the net.

The general idea is using internet anyone can make a survey online using websites such as www.crowdmap.com and using the social community websites such as www.facebook.com or www.twitter.com which most of these websites are free of charge and easy to operate so it will be the most convenient and cost effective to use online surveying for the crowdsourcing. Anyone can simply make an account easily by following the instruction given in the respective website. The data gathered thru Facebook and Twitter are fresh data and we need to plot the graphs or visualize it on a map and timeline but the Crowdmap website will give us the data directly in the form of single platform and visualize it on a map or timeline as we choose.

3. CASE STUDIES

This paper contains 4 case studies which all four cases are comparing their data which gathered thru crowdsourcing with the data which are provided by different government agencies. Each case use different method for gathering information and it is also related and applied in different field or phase of transportation engineering.

The first case is Road Hazards which is basically gather data about road hazards in different areas of National Capital Region and the instrument used in this study is Crowdmap site.

The second case study is Floods in Metro Manila which community observes the water level during rainfalls and by using Crowdmap and Survey Monkey site will share out the results with others. In metro manila one of the main reason for sudden heady traffic during rainy seasons is the flood which will make some roads and streets full of water and impossible for the vehicles to use that route.

The third case is Human-Based Sensory Network for Traffic Volume of Roads which will determine the traffic volume in a specific period along the Epifanio de los Santos Avenue (EDSA) which is one of the major routes in metro manila using crowdsourcing and compare the data with the data of Metro Manila Development Authority (MMDA).

The fourth case is Air Quality Monitoring. One of the most important sources of air pollution in urban areas is vehicles and motors so it has a great impact on the environment. This case use to survey thru crowdmap site and compare the idea of people with the data from government.

3.1 Case 1 (Road Hazards)

One of the main factors that contribute to accidents and traffic congestion nowadays are road hazards. The purpose of this study is to identify the common road hazards which can lead to accidents if ignored. To be able to have an insight and monitor of what most road users observed as they travel from home to office or school through crowd sourcing.

The test instrument employed in this study is the crowd map site. In this site, the respondents will submit reports about their experiences in traffic-related problems such as traffic problems and road hazards. They may also include photos about the traffic condition. The crowd map site consists of questions regarding the respondents' personal information, their title of the report, the description on the traffic condition of problem, choice of category that their report will include and their location where the reported issues happened.

The scope of this research is the observations of the road users and citizens of National Capital Region about the road hazards they encounter as they travel. The study limited itself

to the observations of the road users in the urban areas and the other parts of the country were left out in view. Figure 1. shows the research flowchart which includes the process of gathering data to using the data.

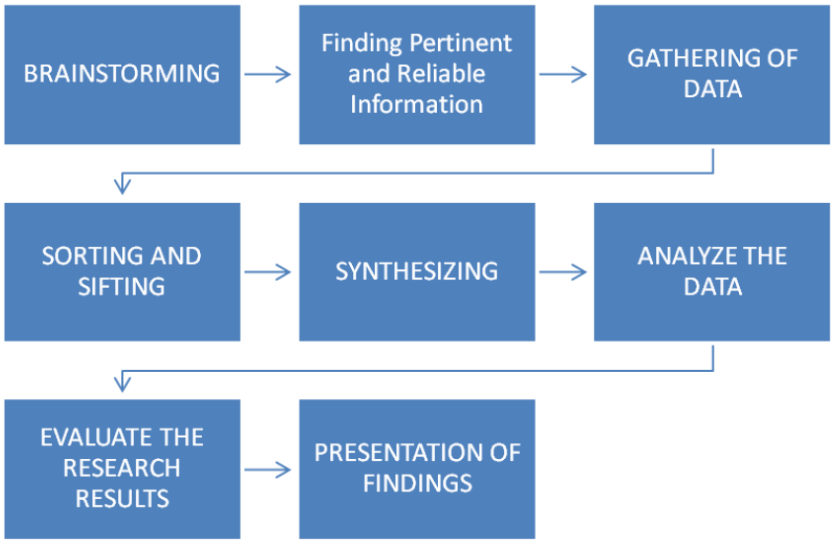

Figure 1. Research flowchart

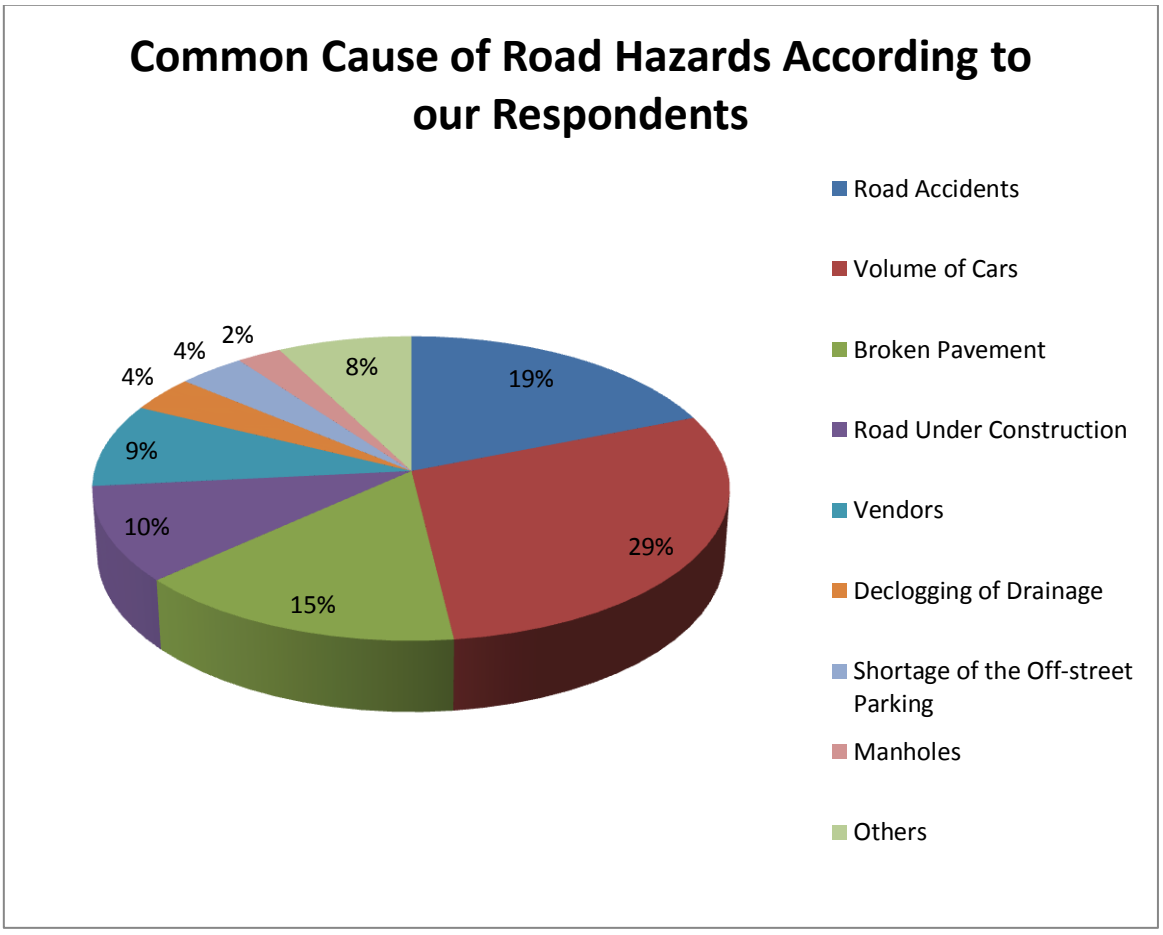

Figure 2. Common encountered road hazards

Figure 2. shows the common problems that manifest on roadways in our everyday lives. This chart shows the common road hazards encountered by the drivers and road users. It

shows that volume of cars have topped the list among other kinds of road hazards according to the results we have garnered.

This case has a total of 100 respondents through crowd map. According to data gathered, common road problems are roadway departure hazards, road surface conditions, narrow roadways and bridges, roadway access and work zones. According to the respondents accident usually happens because of human errors and undisciplined drivers. These types of hazards are usually not noticeable by drivers and local officials.

Table 1. and 2. shows the estimated cost for collecting data of 100 respondents, by doing survey manually and online respectively. Comparing the total cost of survey from 100 respondents, it proves that the cost will be reduced around 60% through online survey comparing it to manual survey. The savings may increase up to 75% as the larger number of respondents is required.

Table 1. Estimated cost of manual survey from 100 respondents

Labor			
	Unit price	No. of days	Total (peso)
Surveyor salary	400 person/day	3 days	1200
Encoding of results	400 Person/day	1 day	400
Materials			
	Unit price	No. of Pcs	Total(peso)
Ball pen	10 peso/piece	3 pcs	30
Photocopy fee	1 peso/piece	100 pcs	100
Total (peso)			1730

Table 2. Estimated cost of online survey from 100 respondents

Labor and Rental			
	Unit price	No. of days	Total (peso)
Salary of online surveyor	500 peso/day	1	500
Rental of computer with internet connection	200 peso/day	1	200
Total (peso)			700

3.2 Case 2 (Floods in Metro Manila)

This study aims to acquire a deeper understanding about the behavior of floods in Metro Manila with the use of crowd sourcing which can provide basis for flood forecasting, especially for riparian communities. By Crowd Sourcing, Flood Hazard Mapping will be designed to increase awareness of the possibility of flooding among the public, local authorities and as well to civilians. This will also encourage societies living and working in flood-prone areas to find out more about the local flood risk and to take appropriate action.

In view of said aim, a survey was conducted amongst residents of Metro Manila. Two websites are used in the study, namely Survey Monkey and Crowdmap, where respondents input some details of the floods they experience in their area. The data of maximum gage height of rivers in Metro Manila was gathered by asking it from the Bureau of Research and Standards, DPWH .The study was limited to certain areas in Metro Manila only.

The results shows that flood mapping can address the areas in Metro Manila where there are risks of floods. The figures below show the efficiency of crowd mapping by comparing the data gathered by the group and the Project Noah of PAGASA.

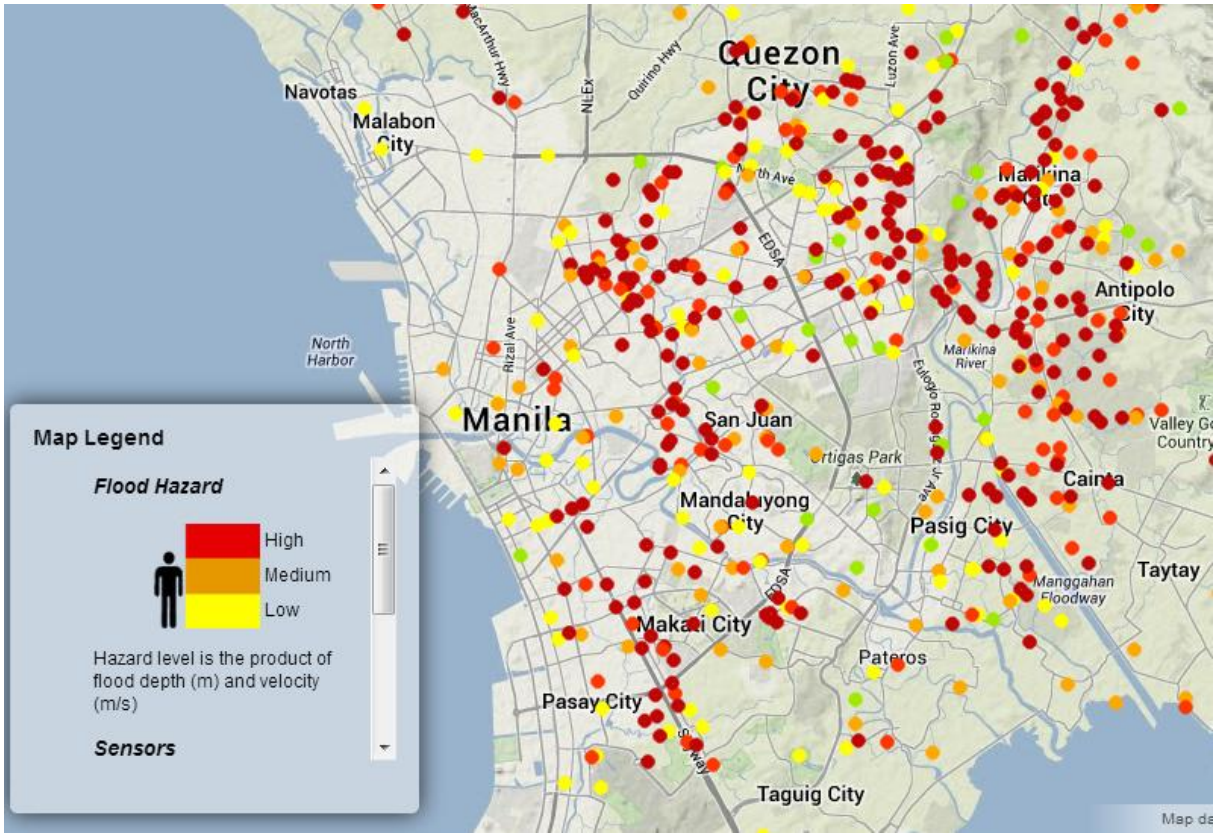

Figure 3. Government data for flood hazard within Metro Manila (project Noah of PAGASA)

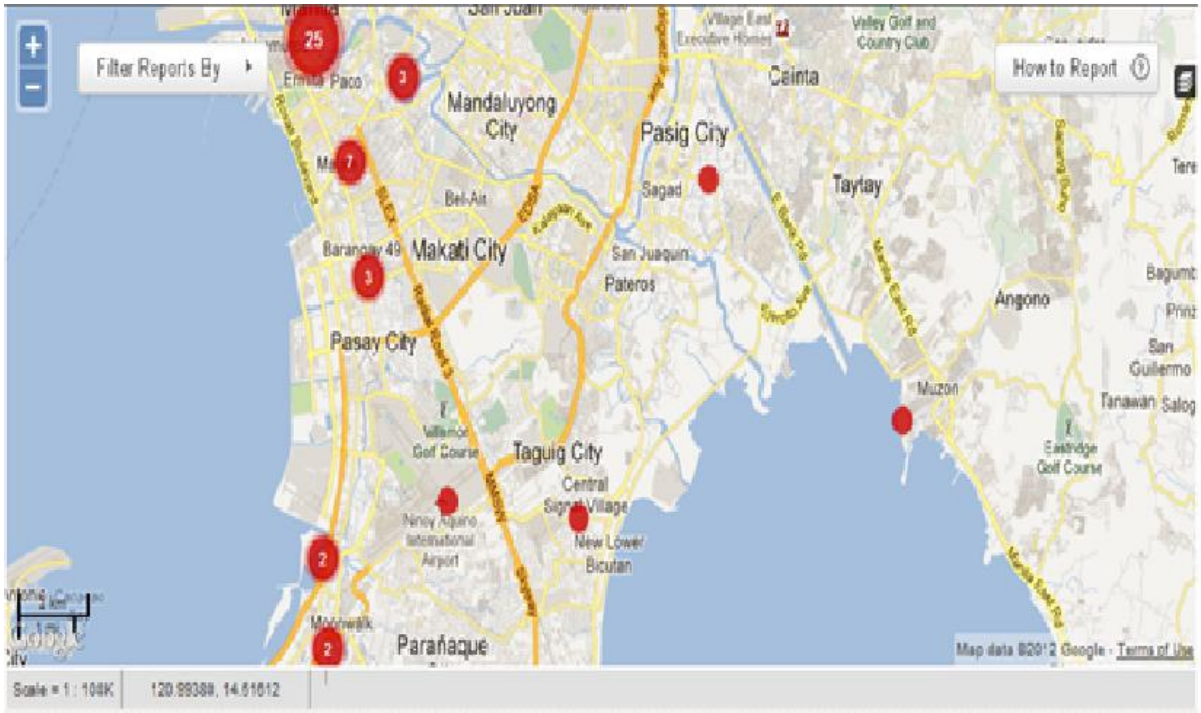

Figure 4. Data gathered through online surveys

The Crowd Map has some similar flood spots with the Project Noah but it does not readily shows the historical flood depth of the area or road. By comparing existing models

and modeling practices and by evaluating them within the same areas rivers and roads, crowd sourcing will be assessed if it is an effective way for flood mitigation.

The findings have shown that such a system of forecasting and response is feasible. The researchers have determined that there are existing projects that encourage and enable civilian participation in recording flood levels and the effects thereof in various communities for usage in fighting off future disasters, and that the water levels reached in times of flooding, especially in riparian areas, have moved people to be more alert, aware, and responsive when it comes to events like these.

3.3 Case 3 (Traffic Volume of Roads)

The epifanio de los santos avenue (EDSA), the 24-kilometer long prime artery of metro manila experience one of the most variable types of traffic. Despite the opening of metro rail transit line 3 (MRT3) and implementation of different traffic schemes like number coding and "yellow bus lanes", EDSA still experiences heavy traffic at certain times of the day. Traffic in this road can be very unpredictable and can change every second. Information about the situation of roads are supplied by the MMDA but is only updated every 10 to 15 minutes. The time duration is too long which means that people might be wasting their time just wait for the next update. This study will be able to identify whether we can rely on crowd sourcing if government agencies are unavailable. Specifically, this study aims to know whether this kind of information sharing over the internet can be used with its advantages shadowing its disadvantages.

To identify what kind of traffic there is in an area, the researchers asked a few people to help gather the data for the research. Those people were asked to identify the kind of traffic the road on their station is experiencing at that time. The data will be obtained through the classification of the traffic at certain times. The classifications were Light, Moderate and Heavy traffic. To classify the current situation, the researchers instructed the people to observe the average car spacing.

Table 3. Average car spacing and traffic condition

Average car spacing	Traffic condition	Level of congestion
More than 2.5 meters	Light traffic	1
In between0.5 to 2.5 meters	Moderate traffic	2
Less than 0.5 meter	Heavy traffic	3

The researchers were able to accommodate five stations along EDSA compared to the 37 stations monitored by the MMDA. The researchers decided to do a 6 hour long survey of the traffic condition along EDSA. Table 4, shows the list of common stations by the researchers and MMDA.

Table 4. List of common stations by the researchers and MMDA

Stations
1. SM Megamall
2. Taft avenue
3. Buendia
4. Santolan
5. North avenue

Table 5, shows the Data which have been gathered from crowdsourcing are almost similar as those which were retrieved from the MMDA.

Table 5. Discrepancy in data obtained

Station, Time and direction	Crowdsourcing (Level of congestion)	MMDA Data (Level of congestion)	Difference Level of congestion
Shaw Blvd., 12NN, Northbound	Light (1)	Moderate (2)	1
Taft Ave., 1PM, Northbound	Heavy (3)	Moderate (2)	1
North Ave., 1PM, southbound	Moderate (2)	Light (1)	1
Santolan, 4PM, Northbound	Heavy (3)	Moderate (2)	1
Buendia, 6PM, Southbound	Light (1)	Moderate (2)	1

As observed, there are slight discrepancies in the observation of the same scene. This can happen when both observers have different quantifiers for traffic. In the case of crowd, they used the car spacing while the MMDA used the amount of cars (volume range). This discrepancy can be neglected since the observations are almost identical. They only time it wouldn't be neglected if difference between level of congestions are more than 1 (e.g. Light (1) for one and Heavy (3) for the other).

3.4 Case 4 (Air Quality Monitoring)

Clean air is rather difficult to come by these days, and this has detrimental effects on human health. Although air quality monitoring in the Philippines has been sporadic and lacks good quality assurance, there is no doubt that the air quality of Metro Manila is seriously degraded.

This case study addresses the air quality of the Metro Manila and Calabarzon through crowd sourcing by developing a website via www.crowdmap.com. Crowd sourcing utilizes an exciting new technology-visibility monitoring; where users can submit reports and even take photos of the sky using their mobile phones, which are then tagged with the date & time, location, and orientation, and submitted to the website, from which they can be analyzed to record air pollution levels for that location.

After the data had been collected, one tabulated the results according to the respondents' specific location, which was divided in two: Metro Manila and Calabarzon. Making tabular results about which area has a greater risk or which area is more polluted. After which, the gathered data was compared to the current data of government, environment and information agencies.

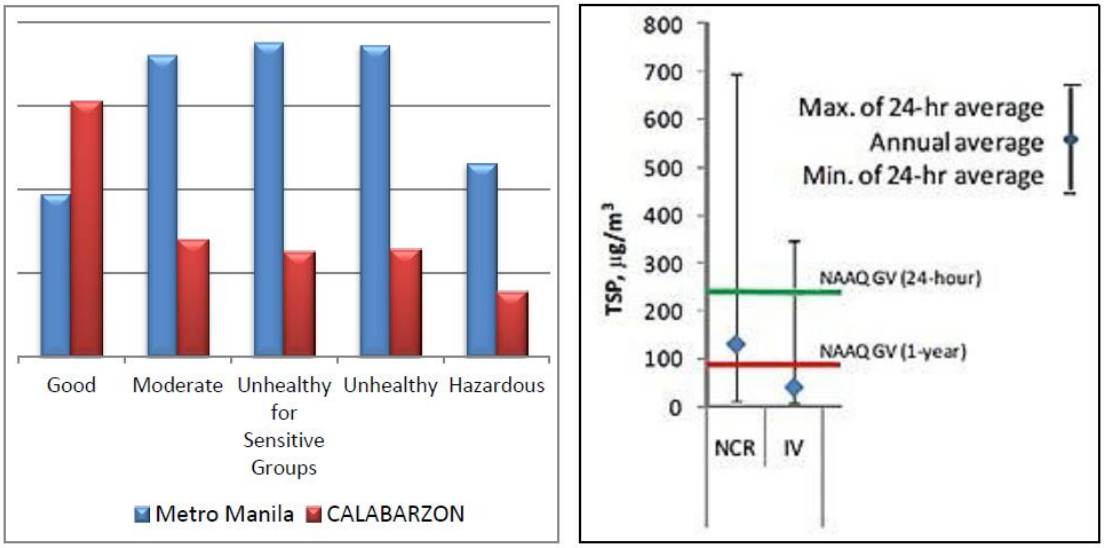

Figure 5. Comparison of the government data and crowd sourcing according to the air quality categories.

Comparing the data from the two locations, the air quality in Metro Manila is polluted than Calabarzon according to the opinion of the respondents. This result matched on the government data, figure 5, on the right, this shows the mean regional annual Total suspended particles (TSP) level, according to the Air Quality Index (for AQI values references proceed to Appendix A), NCR fared fair while Region IV fared good.

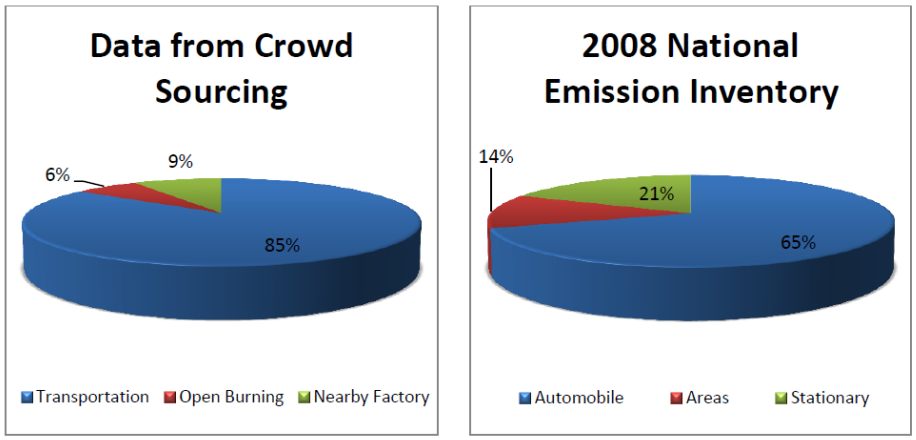

Figure 6. Sources of air pollution in Calabarzon and Metro Manila

Figure 6, on the left shows the possible source of air pollution in Calabarzon and Metro Manila gathered from the survey taken. It shows that majority of the respondents believes that transportation is one of the major sources of air pollution. And least respondents believe that open burning causes air pollution. In the same figure at the right, is a data from 2008 Philippine National Emission Inventory. Based on the 2008 National Emission Inventory, majority of the regions in the country point to the transport sector, consider transportation as the major source of air pollution. As shown in the figure, it was estimated that 21% of the pollutants came from stationary sources or the industry and factory, 65% from automobile sources, and the remaining 14% from area sources, which can be open burning.

This proves that the opinion of the respondents matched with the government data, according to the order of the level.

4. CONCLUSION

In the past few years, the technologies which are related to mobile phones and advanced communication devices have been moving forward, and this issue caused to the most of peoples in all over the world have access to internet connection thru their mobile phones or tablets. These technologies facilitate the use of crowdsourcing in emergency cases like as floods or sudden traffic jam.

As observed in the case studies the data gathered thru crowdsourcing is reliable and it comes from a large community which they willing to inform and help each other in the emergency cases. The most important issue when it comes to the emergency cases is how updated is our data or how often our data will be updated which using crowdsourcing it can be updated each second once or more depending on number of respondent.

One problem which crowdsourcing might be facing is that how to encourage all people to share what they experience or observe. In all cases of this study there was no incentive provided for the respondents so the researchers were having a hard time to find respondents. Encouraging people to participate in crowdsourcing might take a long time but it can happen in near future.

With regards to the method utilized by these cases, crowdsourcing; the idea of soliciting input is hardly new, and the rise of user-generated media movement showed that it can be done with large numbers of people. Crowd sourcing has created a variety of new opportunities for improving upon traditional methods of data collection at a significantly reduced cost.

Crowd sourcing offers an unparalleled benefit, a diversity of thought and experience that simply cannot be obtained any other way than through a crowd with all its diversity; crowd sourcing can be a critical building block of successful innovation.

In quest towards evaluating the risk through crowd sourcing, the researchers conclude that crowdsourcing is very useful tool in today's modern world especially that the populations are rapidly growing and that more innovations are yet to come.

5. RECOMMENDATIONS

One of the most important issues in crowdsourcing method for gathering data is how to encourage people participate in the program. One way to increase the number of active population to participate in crowdsourcing is promotion. We highly recommend promotion through television, social-networking sites and the like.

REFERENCES

Barry, Libert., Jon, Spector. (2010) *Crowdsourcing: A Short Overview for the Rest of Us,* Pearson Education.

Brabham, Daren. (2008) *Crowdsourcing as a Model for Problem Solving An Introduction and Case*, The International Journal of Research into New Media Technologies.

Estellés-Arolas, Enrique., González-Ladrón-de-Guevara, Fernando. (2012) *Towards an Integrated Crowdsourcing Definition*, Journal of Information Science.

Frederic, P. Miller,. Agnes, F. Vandome,. John, McBrewster. (2009) *Crowdsourcing*, Lightning Source Incorporated.
Kiruba, Shankar., Mitchell, Levy. (2011) *#Crowdsourcing Tweet*, THiNKaha, Silicon Valley, California, USA

APPENDIX A

Air quality index (government data)

Pollutant Index		Good	Fair	Unhealthy for Sensitive Groups	Very Unhealthy	Acutely Unhealthy	Emergency
µg/m³	TSP (24-Hr Ave)	0-80	81-230	231-349	350-599	600-899	900-above
µg/m³	PM_{10} (24-Hr Ave)	0 – 54	55 – 154	155 - 254	255 – 354	355 – 424	425 – 504
ppm	SO_2 (24-Hr Ave)	0.000-0.034	0.035-0.144	0.145-0.224	0.225-0.304	0.305-0.604	0.605-0.804
ppm	CO (8-Hr Ave)	0.0 – 4.4	4.5 – 9.4	9.5 – 12.4	12.5 – 15.4	15.5 – 30.4	30.5 – 40.4
ppm	O_3 (8-Hr Ave)	0.000-0.064	0.065-0.084	0.085-0.104	0.105-0.124	0.125-0.374	See Note[1]

Philippine national ambient air quality (NAAQ) guideline values

POLLUTANTS	Short Term µg/Nm³	Short Term ppm	Short Term Averaging Time	Long Term µg/Nm³	Long Term ppm	Long Term Averaging Time
TSP	230	-	24 hours	90		1 year
PM_{10}	150	-	24 hours	60		1 year
Sulfur dioxide	180	0.07	24 hours	80	0.03	1 year
Nitrogen dioxide	150	0.08	24 hours			
Photochemical oxidants as ozone	140 60	0.07 0.03	1hour 8 hours			
Carbon monoxide	35 µg/Nm³ 10 µg/Nm³	30 9	1hour 8 hours			
Lead	1.5	-	3 months	1.0		1 year

Source: ***Philippine Clean Air Act of 1999, section 12***