

PAPER • OPEN ACCESS

Analysis of Water Quality Data in Swimming Waters and Relations with Pollution Sources

To cite this article: Hao Qian 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **283** 012066

View the [article online](#) for updates and enhancements.

Analysis of Water Quality Data in Swimming Waters and Relations with Pollution Sources

Hao Qian

Zhejiang institute of hydraulics & estuary, pr. China

Abstract. In summer, heavy rain will cause water pollution in swimming areas. This study tries to find whether there is potential relationship between peak pollution events and extreme rainfall event. These insights can be used to develop solution to prevent swimming prohibition which caused by peak pollution events.

1. Instruction

In summer, people will crowd to recreational water body to swim for relieving summer heat. Somewhere like beach, canal, pool or lake will be a good choice. However, in The Netherlands, swimming in surface waters is only allowed at dedicated swimming locations. The water quality is checked in regular intervals during swimming season (May to October) at these swimming locations. The warning signs will be put up to forbid swimming at these swimming locations when unacceptable concentrations of toxic substances or pathogens are found at these locations.

The toxic substances mostly relate to microbial pollution, which can occur as a result of different sources, such as combined sewer overflows (CSOs), human feces (especially bathers in recreational area during swimming season), leftovers of barbecue meat, animal (cattle, dog, birds, etc.) dung, etc. The extreme rainfall events are usually responsible for CSOs, so the potential relationships between the extreme rainfall events are started to be concerned. This study tries to find whether there is potential relationship between peak pollution events and extreme rainfall event. These insights can be used to develop solution to prevent swimming prohibition which caused by peak pollution events.

The research question of this study is proposed based on all background information, which is whether there is any potential relationship between extreme rainfall and peak pollution events.

2. Research Method and Available Data

2.1. Statement of Problems and Objects

The pollution of the recreational waterbody is a threat to human health during swimming season. In general, the pollution could be caused by different pollutant sources. Based on the existing data, this research will mainly focus on the pollutions influenced by the extreme rainfall events.

First, a significant amount of pollution entering the sewage treatment plant is therefore from street wash-off during rainstorms. Unless the combined flow of sanitary sewage and storm water exceeds the interceptor capacity, it is captured by a regulator, and then transferred to the wastewater treatment plants. If the combined flow exceeds the interceptor capacity, however, overflows are directly discharged to a receiving river, and these are called combined sewer overflows (CSOs) (Kim, 2007). Because CSOs are composed of a mixture of surface runoff, sewer deposits, and untreated sanitary sewer



flow, they can contain various pollutants of very high concentrations, especially at the early phase of a rainfall event (Lape and Dwyer, 1994). It is widely assumed that the frequency or total volume of overflow discharges is a good indicator of the pollution impact on receiving waters. The less frequently an overflow spills, and the lower will be its negative impact on receiving water quality (James, et al., 2002).

In about sixty percent of the Netherlands surface area, there is no natural drainage. The surface water and groundwater levels are controlled by a complex network of canals and pumping works. The hydraulic head differences in sewerage systems have to be created almost entirely by sewage pumping stations. The heavy rainfall event will cause the failure occurring since the limited capacity of pumping stations which will lead overflows. Intensive rain can also cause re-suspension of deposits, generally resulting in increased pollution concentrations in overflow discharges. The combined sewer overflows cause great disturbance of the receiving water environment. In some systems, a significant feature is the first flush (David, 2004) in early storm flows, which may contain particularly high pollutant loads. An obvious sign would be a sharp increase in pollutant concentration near the start of a storm.

Second, rainfall leads to stormwater runoff which reaches some areas does not receive any pre-treatment. Stormwater runoff on streets draining residential areas has been found to contain high concentrations of fecal coliform bacteria (Bannerman et al., 1993; Young et al., 1999). Sources of fecal bacterial pollution entrained in runoff from impervious surfaces in urban and suburban areas are rodents and other small mammals, avifauna and especially pets (Mallin, et al., 2001). Numerous manure deposits (especially from dogs) on sidewalks, roadsides, tree lawns and around apartment dwellings are often within meters of storm drains. The human feces, left overs of barbecue meat are also considered as the pollution sources (Cransberg, et al., 1996). The general stormwater runoff, particularly from impervious surface areas, is the major source of microbial pathogen pollution.

The object of this research is by analyzing the existing data to find potential relationships between extreme rainfall events and peak pollution events of swimming water. This information can be used to develop solution to prevent swimming prohibition which caused by peak pollution events.

2.2. Available Data

For this project, we have precipitation data for every 10 minutes from 2008 to 2011 and swimming water quality seasonal reports from 2008 to 2011 for Rijnland area which are prepared by water authorities (Delfland, Rijnland, provinces of Zuid-Holland and North-Holland and others). Since the water quality data is measured around 12 times (mainly from April to September) one year, so the time line is rather different from the precipitation data.

The rainfall data for study area 'Rijnland' is accurate to every 10 minutes from 7 rainfall observation stations, which are marked in Figure 1. The red and yellow spots are sampling points for swimming water quality.

The extreme rainfall could cause the combined sewer system overflows. According to the background information, it indicates the pumping station is slowly emptying the system during storm event. The design criterion which was traditionally used in the NL is: 9 mm storage + 0.7mm/h pumping capacity. So it defines if any storm intensity near or above this rainfall amount is likely to cause an overflow.

For the swimming water quality data, the water samples have two different measuring units, which are *E. coli* (MWA/100ML 2008 - 1010) and Intestinal Enterococci (KVE/ML for 2008 – 2009; MWA/100ML for 2010 - 2011). The measuring period for the swimming water quality is normally from April to September, the sampling time every month is not exactly the same. The water quality data is measured around 12 times per year. There are 51 sampling points (Figure 1) in total.

2.3. Method

In order to find the potential relationship between the extreme rainfall event and the extreme pollution incident, the data which represent the extreme rainfall should be defined. At the beginning, the average

precipitation of whole area is considered. However, after processing the existing data, the results show that there are rather big differences of precipitation data between different observation stations. So the average rainfall data of the whole area cannot be used to do the comparison. This implies that a responsible boundary for each rainfall observation station should be defined. Afterwards, the precipitation data for each observation station (for example: station X) can be used to do the comparison with the swimming water quality data from these sampling points which within station X's responsible boundary.

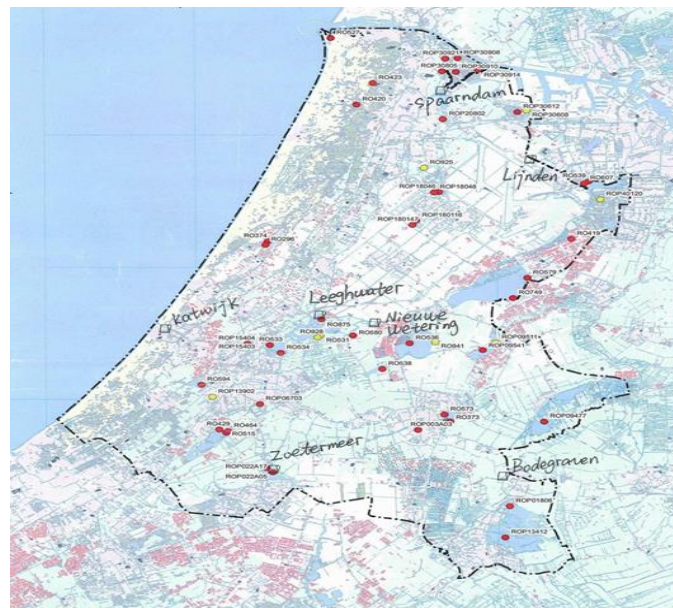


Figure 1. Study area Rijnland with 7 rainfall observation stations and 51 swimming water quality sampling points

The study area for this project could be defined as a plain river network, if we look into the whole study area as a large-scale. Comparing to the mountain area, the slope differences of plain river network are not so huge. Therefore, if this area is equally divided into several small zones, the transfer of water could keep a balance within their own zone. For this reason, it is proposed to divide our study area equally to 7 zones base on the 7 rainfall observation stations (Figure 2). To achieve this, each rainfall observation station is assigned approximately the same area (the green line is the boundary of each responsible area, and the rainfall stations are in pink circle). Then, the comparisons between the precipitation data from one rainfall station and the pollution data from these sampling points within this rainfall station's responsible boundary can be done.

By dividing the study area into 7 responsible boundaries:

Area 1 - Rainfall station Spaarndam: ROP30921, ROP30908, ROP30805, ROP30910, ROP30914, ROP20802, ROP30612, ROP30608, RO423, RO420, RO527;

Area 2 - Rainfall station Lijnden: ROP18046, ROP18048, ROP180116, ROP180147, ROP40120, RO539, RO607, RO419, RO925;

Area 3 - Rainfall station Leeghwater: RO374, RO296, RO875, RO928, RO531, RO580;

Area 4 - Rainfall station Nieuwe Wetering: ROP09511, ROP09541, ROP09477, RO536, RO538, RO941, RO579, RO749;

Area 5 - Rainfall station Katwijk: ROP15404, ROP15403, ROP13902, RO533, RO534, RO594;

Area 6 - Rainfall station Zoetermeer: ROP06703, ROP022A17, ROP022A05, RO429, RO464, RO515;

Area 7 - Rainfall station Bodegraven: ROP003A03, ROP01806, ROP13412, RO573, RO373

After defined responsible boundaries, it is possible to start analysis the existing data within each boundary.

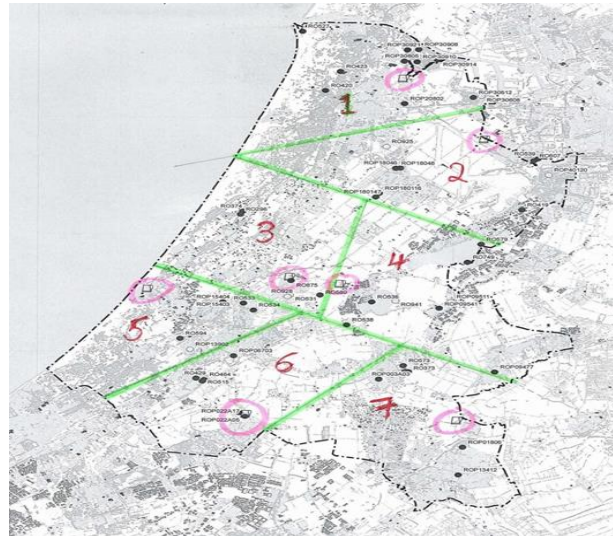


Figure 2. The study area is divided into 7 responsible boundaries based on 7 rainfall stations

3. Analysis steps

3.1. The First Try Out

At the beginning of the analysis, the extreme values which can represent the extreme rainfall events are defined as: if rainfall intensity is keeping more than 2mm/10min, the overflow will happen. So the value '2mm/10min' is defined as the critical value to represent the extreme rainfall event. In order to find the potential relation between extreme rainfall event and extreme pollution event, the first idea is to compare the extreme values directly. However, it is hard to compare these two groups of data by using Excel, since the time line for each group is so different. It is tried to draw all extreme rainfall points and pollution points manually, but the results did not show clear relation and make no sense at all.

3.2. The Second Try Out

For this try out, the software MATLAB is used to process data. All precipitation data and pollution data is shown in figures by time series. Since the pollutant has two different units (MWA/100ML and KVE/ML) for 2008 - 2009, and the same unit (MWA/100ML) for 2010 - 2011, so each study area has 6 comparison figures from 2008 to 2011. There are 42 figures in total for the whole study area. For the first figure (Figure 3), the original idea is attempting to figure out whether the extreme rainfall events will be followed high pollution spots. However, the results also cannot show clear relation between rainfall data and pollution data, especially for extreme event. It implies that to compare these two groups of data by time series is not a good solution. In order to find the potential relation between the extreme rainfall event and the peak pollutant, it is better to compare these two factors directly.

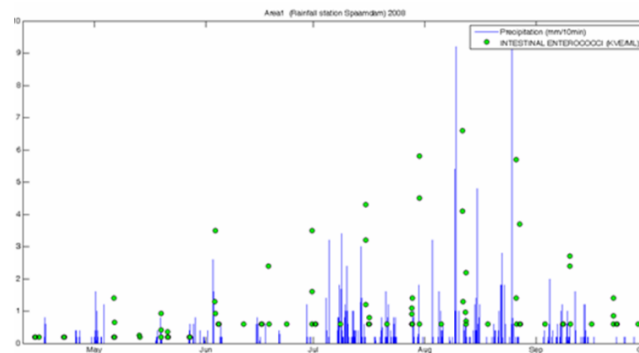


Figure 3. The results of rainfall and pollutant for area1 (2008) in the second try out analysis

3.3. The Third Try Out

In order to find the answer to research question, the characteristic data which represent extreme events should be defined. As the extreme rainfall data is defined before (2mm/10min), it seems here just need to find pollution data to compare with the precipitation data. However, the pollution data sampling frequency is not according to the extreme rainfall event in this case. So if using extreme rainfall event to compare with the pollutant which is sampled several days later (it is hard to define this pollution is caused by the extreme rainfall days before), it is not make sense.

The exiting data show that the rainfall data is sufficient (every 10 minutes for 4 years), but the pollutant data is limited (around 12 times per year). In order to find the relationship between extreme rainfall event and pollutant, it is better to make full use of pollution data. It is considered to do the analysis start at the pollutant data.

In order to indicate whether the extreme rainfall event will cause high pollution, it is need to know the rainfall data just before the pollution sampling date. For this try out, limiting by poor programming in MATLAB, it just filtered out average rainfall of one week just before every pollutant sampling date. The rough comparisons are made between weekly average rainfall just before the sampling data and the pollution data, which means to get the trend of a general picture of the relation between rainfall and pollutant (Figure 4). This figure represents the results of the comparison between pollutant (MWA/100ML as the Y axis) and weekly average rainfall before the pollutant (mm/10min as the X axis). It assumes the weekly average rainfall before pollutant sampling can approximate represent the behavior of extreme rainfall event. If there is extreme rainfall event within the week before pollutant sampling date, the average rainfall data will also show a rather extreme behavior. Later on, this assumption is being proved not true.

The limited pollution data is also the main problem for this try out, since it is hard to define extreme rainfall events just before each pollution data. The results still did not match the research question very well, since the weekly average rainfall depends very much on the total of rainfall events in this week. The results of the third try out will not be taken into account in the next step.

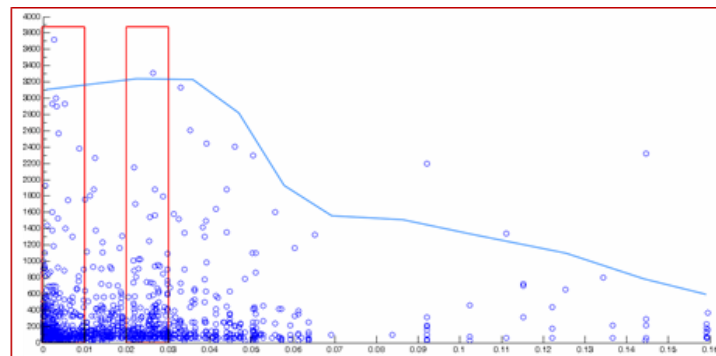


Figure 4. Integration results for the comparison between pollutant (MWA/100ML) and weekly average rainfall before the pollutant (mm/10min)

3.4. The Fourth Try Out

By re-programming in MATLAB, strong boundary conditions are set for data processing. For the first processing, it defines that if:

- ① The rainfall intensity is ≥ 2 mm/10min for more than 20 minutes;
- ② Condition ① is repeated more than 2 times;
- ③ Take the rainfall data 2 days before the pollution data into account.

If the rainfall and pollution data match all conditions above, then this group of results will show in figures. Otherwise, the results are ignored. After processing existing data within such conditions, most results show error. This is because these conditions are too strong for such limited pollution data.

Therefore, for the second processing, the condition ② is changed to : if condition ① is repeated more than once. After processing and filtering existing data, there are 29 groups of valid data show in the figures, which seem like reasonable results for this case. The average extreme rainfall (filtered out by the program in MATLAB which means it possibly represent an extreme rainfall event) two days before the sampling date is used as X axis. The pollutant concentration is used as Y-axis (Figure 5).

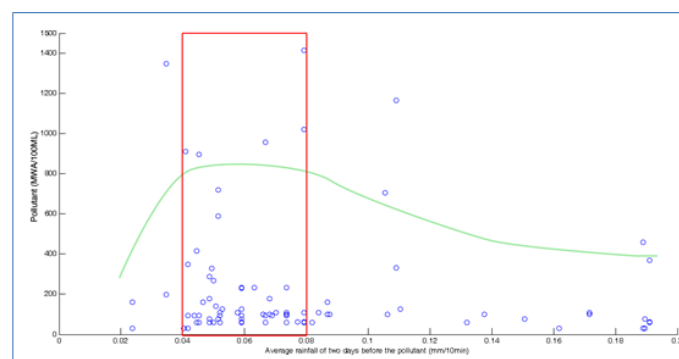


Figure 5. Integration results for the comparison between pollutant (MWA/100ML) and average rainfall (filtered out by MATLAB which possibly represent extreme rainfall) 2 days before the pollutant (mm/10min)

Later on, there are two problems been questioned in the preview analysis. The first is the definition of the extreme rainfall event (2mm/10min), this extreme values are not strong enough to represent an extreme rainfall event. The second problem is using the average extreme rainfall intensity in the comparative figure is not a good choice also.

3.5. *The Final Try Out*

According to all previous try out analysis and experience. For this analysis, the boundary condition using for programming in MATLAB is even stronger. It considered the traditionally design criteria used in The Netherlands, which is 9 mm storage + 0.7mm/h pumping capacity. This means any storm near or above this rainfall amount is likely to cause an overflow.

For processing the data and filtering out the possible overflow event, the program is set to compute the inflow, outflow and also the real-time storage with the precipitation data one day before the pollutant data (with the design criteria 9 mm storage + 0.7mm/h pumping capacity). If the cumulative rainfall amount is more than the storage (possible overflow events), this group of data will be filtered out and put into comparative figures. The pollutant concentration is used as Y axis and rainfall intensity (filter out by programming in MATLAB which represent an overflow event) one day before the pollution date is used as X axis. The program is being processed 42 times for each area in total, within different years and in different pollutant units. Since the stronger boundary condition, this time there is 25 groups of comparison been filtered out by processing in MATLAB, and the pollutant spots in each comparative figure are also much fewer. The detailed results could be found in following chapter.

4. Results and Discussion

According to previous research, it shows extreme rainfall will cause combined sewer system overflow which will lead a peak pollution event in swimming water-body. Also human feces (especially bathers in recreational area during swimming season), leftovers of barbecue meat, and animal dung from street also will be flushed into nearby recreational water-body.

Before data processing in the final try out, the hypothesis is like the third and fourth try out. Based on the research question “whether there is any potential relationship between extreme rainfall and peak pollution event”, it assumes that the pollutant concentration will increase as the rainfall intensity (causing overflow event) before the sampling date is increasing, or the pollutant concentration will increase first and decrease later.

After processed the existing data under the final try out program in MATLAB, there are 25 groups of valid comparisons. However, the valid pollutant point show in each figure is very few because of the stronger boundary conditions (base on design criteria: 9 mm storage + 0.7mm/h pumping capacity). Therefore, it is hard to indicate any potential relation between rainfall intensity (causing overflow event) and pollutant from each single figure.

Examples of valid comparisons are shown as below:

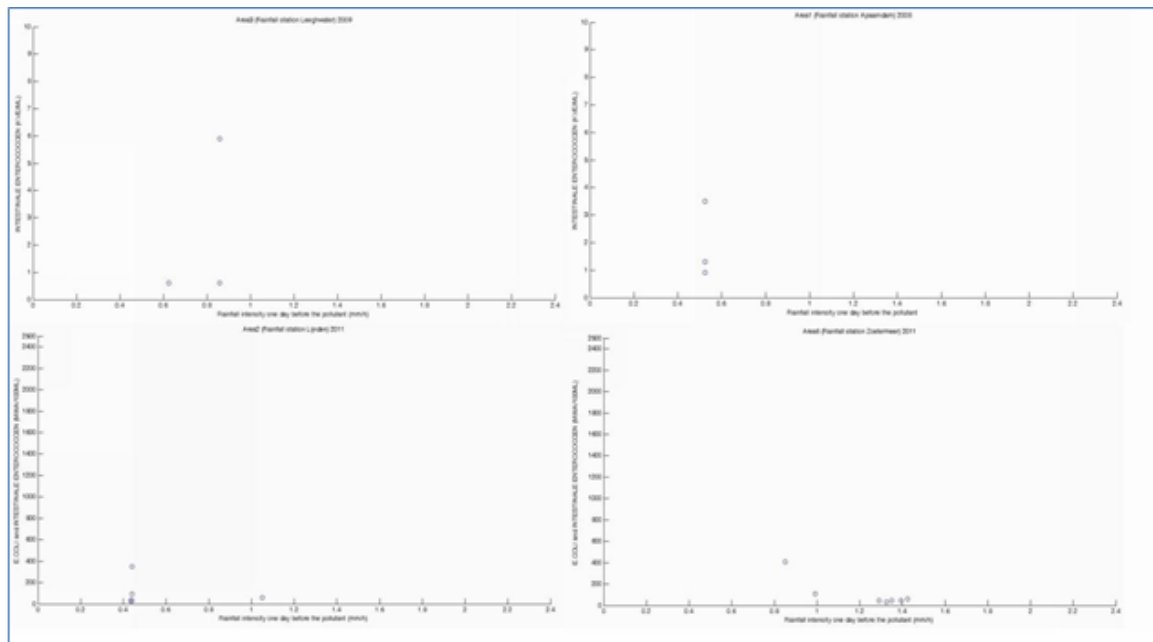


Figure 6. Examples of valid comparisons after processing and filtering

In order to find more clear relation from these results, it is better to put results together based on two different pollutant units. The integration of the results is shown as below:

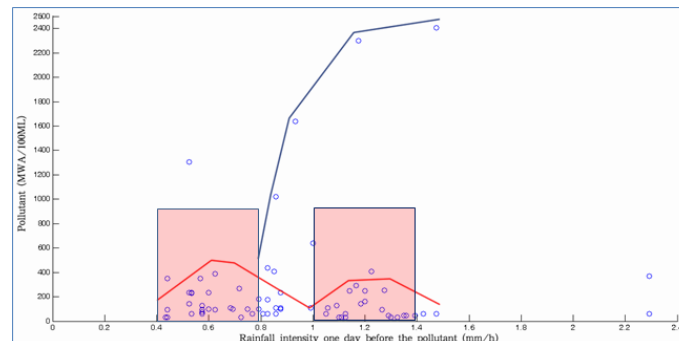


Figure 7. The integration results for comparison between pollutant (MWA/100ML) and rainfall intensity (causing overflow event) one day before pollutant

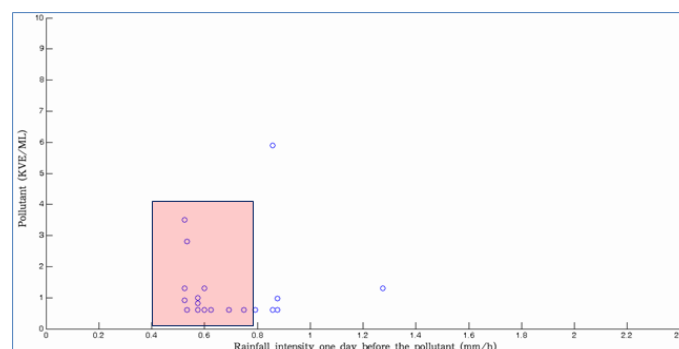


Figure 8. The integration results for comparison between pollutant (KVE/ML) and rainfall intensity (causing overflow event) one day before pollutant

The integrated results from figure 7 and figure 8 show some general relationship between extreme rainfall event and peak pollutions, especially figure 7. However, this relationship is not so clear and strong, because of the limited pollutant sampling frequency. The potential relationship between extreme rainfall event and peak pollutions from the analysis results can be concluded as:

(1) In Figure 7, the pollutant (MWA/100ML) spots are concentrated in two regions (in red box), where the rainfall intensity is 0.4 mm/h – 0.8 mm/h and 1 mm/h – 1.4 mm/h;

(2) Within these two regions, the relationship between pollutant and rainfall intensity show a similar trend: the pollutant concentration increased first as the rainfall intensity (one day before the sampling time and causing overflow event) increased and decreased later;

(3) In region 0.8 mm/h - 1.6 mm/h, there are 4 extreme high pollutant spots (as the blue line shown in Figure 7). If these 4 data are not measuring error and take them into account, it implies that the peak pollution event increased as the extreme rainfall amount increased. This matches the first hypothesis;

(4) After rainfall intensity 1.6 mm/h, there is almost no pollution spots appeared anymore except two spots show in region 2.2 mm/h – 2.4 mm/h;

(5) In Figure 8, the valid pollutant (KVE/ML) spots related to extreme rainfall event (causing overflow) are very few. The pollutant spots are all concentrated in region 0.4 mm/h – 0.8 mm/h (in red box). Some high pollution spots appeared around rainfall intensity is 0.5mm/h.

5. Conclusions

By processing and analyzing with existing data, the results did not show a strong potential relationship between extreme rainfall event and peak pollution. However, by integrating these processed and filtered results, it shows:

(1) For pollutant (MWA/100ML), the valid pollution spots are concentrated within extreme rainfall region 0.4 mm/h – 0.8 mm/h and 1 mm/h – 1.4 mm/h. Four extreme high pollution spots appeared in extreme rainfall region 0.8 mm/h – 1.6 mm/h, which increased as the extreme rainfall amount increased. After rainfall intensity 1.6 mm/h, there is almost no pollution spots appeared anymore except two spots show in region 2.2 mm/h – 2.4 mm/h.

(2) For pollutant (KVE/ML), the valid pollution spots are concentrated within extreme rainfall region 0.4 mm/h – 0.8 mm/h. and some high pollution spots appeared around rainfall intensity is 0.5mm/h.

For this project, the pollutant sampling frequency is rather low to detect a very clear potential effect of extreme rainfall. Also, in order to go deeper in this research, it is not enough to explain why the peaks occur at certain periods in time. The weather conditions and available local data of water system are also need to be taken into account, then to develop solution to prevent peak swimming water pollution events.

References

- [1] Bannerman, R.T., D. W. Owens, R. B. Dodds & N. J. Hornewer, "Sources of pollutions in Wisconsin stormwater", *Wat. Sci. Technol.* 28, 1993, pp 241-259.
- [2] David Butler, John W. Davies, "Urban Drainage - second edition", 2004, pp 256-288.
- [3] Geonha Kim, Joonghyun Yur, Jeongkon Kim, "Diffuse pollution loading from urban stormwater runoff in Daejeon city", *Journal of Environmental Management* 85, 2007, pp 9–16.
- [4] James Lau, David Butler, Manfred Schutze, "Is combined sewer overflow spill frequency/volume a good indicator of receiving water quality impact? ", *Urban Water* 4, 2002, pp 181–189.
- [5] K. Cransberg, J.H.C.T. van den Kerkhof, J. R. J. Banffer, C. Stijnen, K. Wernars, N. C. A. J. van de Kar, J. Nauta and E. D. Wolff, "Four cases of hemolytic uremic syndrome – source contaminated swimming water", *Clinical Nephrology*, vol. 46, January 1996, pp 45-49.
- [6] Lape, J., Dwyer, T.J., "Controlling combined sewer overflows (CSOs): EPA's CSO control policy", *Proceeding of the 1994 WERF, Kentucky, TN*, 1994, pp 1–18.

- [7] Michael A. Mallin, Scott H. Ensign, Matthew R. McIver, G. Christopher Shank & Patricia K. Fowler, "Demographic, landscape, and meteorological factors controlling the microbial pollution of coastal waters ", 2001, pp 185-192.
- [8] Young, K. D. & E. L. Thackston, "Housing density and bacterial loading in urban streams", *J. Env. Eng.* 125, 1999, pp 1177-1180.