

# Mathematically Aided Risk Assessment of Crude Oil Contamination in Ogoni, Nigeria

## Part 2. Generalizing Statistical Analysis of the Pollution

By H. THIERGÄRTNER and K. HOLTZMANN\*

### Abstract

Mathematical modeling can support environmental risk assessment and decision making processes. Soil contamination caused by crude oil in the Ogoni region, Nigeria, is qualitatively described in detail in part 1 [2] of this three-part contribution to understand the presented mathematical results. A mathematical-statistical analysis following in part 2 characterizes quantitatively 33 contaminated sites as entire ecological complex. The sites are studied in part 3 by classifying multivariate mathematical models to derive precise information about kind and degree of contamination at every single oil spill.

The data basis for the environmental risk assessment supported by part 2 consists of about 660 samples analyzed for UNEP. It is restricted to aliphatic and aromatic hydrocarbons.

Data have been interpreted first using generalizing mathematical-statistical analyses. The entire data set was understood as random sample in the statistical sense. So far, all results are valid regarding the contamination of all investigated sites in Ogoni as a whole unit regardless of local specifics.

Concentration values of aliphatic and aromatic hydrocarbons in soil samples show a statistically significant lognormal frequency distribution: most soil samples are more or less slightly contaminated but extremely high polluted points and sites occur also frequently. Based on the distribution type, the expected (averaged) value was determined as about 3500 mg/kg aliphatics and 1700 mg/kg aromatics. Aliphatics and aromatics are jointly strong and linear correlated. This shows that there is no natural differentiation between the main contaminants caused by migration etc. The degree of contamination is not related to the sampling depth. This unexpected result will be explained by previous, incomplete clean-up measures.

The general statistical evaluation defines dimension and problems of future remediation measures but it is insufficient to determine kind and degree of pollution at specific pol-

luted sites. Only classifying models are suitable to solve this problem.

### Data Basis

The Ogoni region in Nigeria is known as an important source of crude oil. Exploration and exploitation of the deposits and transport of the crude oil as well led to numerous oil spills endangering essential parts of the natural ecology and the basis of human life [2]. Previous attempts to clean-up heavily contaminated sites have not been completed. 33 sites determined by UNEP which are most likely contaminated by crude oil or kerosene have been selected for an environmental risk assessment supported by mathematical models. An exact assignment to local names of locations is not always possible without doubt because of a negligent description of the soil sampling locations and the soil samples itself. Another reason for the difficult assignment of sampling locations to local names is the often doubtful and controversial localization of sites to specific Ogoni communities.

Overall 280 shallow wells were drilled within this area by UNEP contractors in 2010 [3]. Thus, eight to nine sampling locations can be attached to every investigated site.

Altogether 665 interval samples have been taken from these bore-holes covering the total drilling depth. Each bore-hole yielded two to three samples averagely. The intervals are reflected for the mathematical processing by their average depth. The distribution of sampling depth (Tab. 1) shows that a great amount of samples is located within the uppermost soil layers. A representative evaluation

of deeper subsoil layers is therefore more difficult.

The average sampling depth calculated from 804 soil samples is  $92.8 \pm 69.6$  cm.

Chemical analyses regarding all samples are available only for the attributes aliphatic and aromatic hydrocarbons respectively as sum parameters. However, there are many constituents in crude oil which are eco-toxicologically dangerous. On the other hand, aliphatics and aromatics are sufficient to explain mathematical models and their results. All analytical data were taken from UNEP reports and could not be checked with respect to their accuracy and precision. It is assumed that they are faultless and correct.

The data set was not prepared at first for mathematical processing. Some transformations had to be made to create a computable data file: (1) Less-than-values as  $<10$  mg/kg or  $<100$  mg/kg were replaced by 5 and 50 mg/kg respectively to allow a mathematical approach. (2) Data were rounded to tens of mg/kg to simplify an interpretation. (3) Missing values caused an elimination of the corresponding sample even if they were denoted by (not measured!) zero values. (4) True zero values (0 mg/kg) were transformed into ones (1 mg/kg) to avoid errors at logarithmic transformations.

Results of a mathematical processing shall be compared with target or intervention values listed in national or international guidelines. The maximum tolerable value of the concentration of total petroleum hydrocarbons (TPH) in the Nigerian practice is 5000 mg/kg. This value alone is insufficient to assess a chemically complex contamination. Target and/or intervention values are necessary for other compounds which are

Table 1 Distribution frequency of sampling depths

Depth (cm)	Mean (cm)	Number of samples	Percentage of samples
0–20	10	249	31
20–50	35	137	17
50–100	75	130	16
100–150	125	155	19
150–200	175	75	9
200–250	225	44	5
250–300	275	17	2

Table 2 Soil target and intervention values [1]

Contaminant	Target value (mg/kg)	Intervention value (mg/kg)
Mineral oil alkanes	50	5000
Benzene	0.05	1
Ethylbenzene	0.05	50
Toluene	0.05	130
Xylenes	0.05	25
Phenol s. s.	0.05	40
Sum of polycyclic aromatic hydrocarbons	1	40

\* Prof. Dr. Hannes Thiergärtner, Faculty of Geosciences, Free University Berlin, Germany (thiergartner@aol.com); Dipl.-Geol. Kay Holtzmann, Consulting Geologist, Berlin, Germany (kay-holtzmann@t-online.de). Part 1 has been published in the last issue, part 3 will be published in the following issue.

ecologically hazardous as well. Limiting intervention values [1] (Tab. 2) will be applied alternatively to discuss obtained mathematical results as follows: 5000 mg/kg for aliphatic hydrocarbons and 100 mg/kg for the sum of aromatic hydrocarbons.

### General Statistical Description

**Topic.** A general statistical description is a helpful tool to create an objective and reproducible view over the data set independent of the investigator's conceptions. Even if it does not solve a concrete local problem, it is suitable to derive following steps of investigation. It is impossible to analyze and to describe the environmental state of the total Ogoni area, of every point and each site of it. On the other hand, it is necessary to obtain reliable information about a surveyed site and its possible quality and degree of contamination.

**Methods.** The problem can be solved by stochastic modeling. Although the true state of contamination in the entire region is unknown in principal, it can be "estimated" – that is calculated – from a representative set of randomly selected samples and their chemical analyses. Information obtained from this set will be notionally extended to the whole studied object with a chosen confidence interval, e. g. 90% or 95% certainty. The set of single (physical) samples will be understood as one "random sample" in a statistical sense. A random sample is a set of physical samples (specimen) or values randomly taken. "Randomly" means: without any reference to the target of investigation or the studied attributes respectively. For instance, physical samples taken with respect to a grid pattern are randomly obtained. This assumption allows an unbiased statistical treatment.

811 chemically analyzed soil samples are available. Among them, 665 were analyzed for aliphatics and 679 for aromatics; thus a common evaluation of aliphatic and aromatic concentration values can be accounted only for 665 samples. They form the random sample. All results obtained from them can be stated as valid for all of the 33 sampled sites in Ogoni.

The statistical processing is composed (1) of description, analysis and discussion of the frequency distribution of each considered contaminant, (2) of the calculation of compatible average values for the entire data set, (3) of the analysis of mutual relations between all involved parameters, and (4) of the investigation of any dependencies of contaminants on the sampling depth. All results will characterize the data set representing the entity of more or less contaminated subsoil samples and sites in Ogoni as a whole. Mathematical-statistical information can be applied to create an overview but not to generate statements with respect to single soil samples or to certain, possibly polluted sites.

### Distribution of the Degree of Contamination

**Topic.** The distribution of a pollutant describes best in which manner the manifold of measured data covers the concentration axis from the minimum to the maximum value. It is an univariate attribute and has to be determined for every involved parameter.

**Methods.** A simple but clear method is the graphical representation. All values will be classified to given groups and the frequency of occurrence will be displayed as a histogram. A less subjective mathematical method is an adequate statistical test, e. g. the  $\chi^2$  test or the Kolmogorov-Smirnov test. Applying proper software, these tests can be carried out by computer. The algorithms are proving whether the observed frequencies of occurrence can be fitted to one of well known theoretical distribution functions or not, based on a given statistical significance (e. g. 95%).

**Results.** The histograms created for aliphatic and for aromatic hydrocarbon contents in soil show typical L-shaped curves. They are asymmetrical or skew to the left: Low concentrations of contaminants are predominating (Fig. 1), although also numerous large concentration values occur. It is obvious that not all studied sampling locations and not all sites show dangerously high pollution whereas other ones require remediation measures. Tables 3 and 4 complete the graphics and show the percentages of occurrence for classified concentrations. Thus, 91.4% of all analyzed samples are polluted by aliphatic hydrocarbons below the intervention value and the corresponding number for aromatics is 52.6%. Both data allow a first evaluation about the dimension of possible revitalization measures.

The unbiased statistical test yields lognormal distribution types for both contaminants. Lognormal distribution means that the logarithms of concentration values can be approximated by Gaussian distribution curves (Fig. 1). The concentration values of many mineral deposits or polluted sites as well show lognormal distributions.

### Average Impact

**Topic.** The average value of contamination calculated for all analyzed samples characterizes the investigated sites as a whole.

**Method.** Usually, (arithmetic) mean values will be calculated for every compound. It is less known that arithmetic mean values are correct ones only in case of normal or Gaussian distributions. Every distribution function is characterized by its own algorithm (formula) to calculate the true and unbiased mean value or standard deviation as well. The above obtained knowledge of the present distribution type – lognormal distribution – consequently allows applying the right formulas to estimate the unknown but true expected value  $\mu$  (mean value) and standard deviation  $\sigma$  (dispersion) if needed.

**Result.** The statistical mean values  $m_{ig}$  as estimators of the expected values are 3510 mg/kg for aliphatics and 1690 mg/kg for aromatics. As an average, the impact by aliphatics is below the intervention value applied in Nigeria but the concentration of aromatics in subsoil is unacceptably high and above all international standards. Both mean values characterize the contamination observed at all 33 sampled sites but not the Ogoni region as a whole.

By the way, both average values are remarkably higher compared with the corresponding and usually calculated arithmetic mean values (Tab. 5). This is caused by the decided skewness of the distribution function (or curve) and by its great standard deviation. Nevertheless, it is inaccurate to introduce

Table 3 Summarized distribution of contaminants

Aliphatics		
Contamination (mg/kg)	$n_i$	%
0–50	232	34.9
50–100	53	8.0
100–200	62	9.3
200–500	84	12.6
500–1000	56	8.4
1000–2000	51	7.6
2000–5000	70	10.6
5000–6000	20	3.0
6000–7000	6	0.9
7000–8000	7	1.1
8000–9000	5	0.7
9000–10000	0	0.0
Aromatics		
Contamination (mg/kg)	$n_i$	%
0–20	207	30.5
20–50	84	12.4
50–100	66	9.7
100–200	66	9.7
200–500	89	13.1
500–1000	56	8.3
1000–1500	30	4.4
1500–2000	25	3.7
2000–3000	17	2.5
3000–4000	14	2.0
4000–5000	9	1.3

Table 4 Percentiles of occurrence of concentration classes

Percentile	Concentration in mg/kg	
	Aliphatics	Aromatics
10	10	0
20	20	10
30	40	20
40	90	50
50	170	90
60	350	190
70	732	350
80	1872	780
90	4178	1820
100	28,200	13,800
Number of involved samples	665	679

the lower arithmetic mean values instead of the correct lognormal ones e. g. to “diminish the problem”. The statistical analysis of the frequency distribution shows unmistakably that the “traditionally” calculated “mean” concentrations of 1360 mg/kg aliphatic contamination and 620 mg/kg aromatic pollution must not be applied in this case.

### Association between Aliphatic and Aromatic Hydrocarbons

*Topic.* It is important to be informed about all relations between attributes involved in the investigation to decide on number and kind of pollutants that must be studied in detail. In general, mutual relations between components can vary in crude oil depending on its source.

*Method.* (Univariate-statistical) correlation analyses are the proper tool to study existence and degree of mutual dependencies between attributes. The corresponding measure is the correlation coefficient  $r$ .  $r = 0$  indicates absence of correlative relations. A coefficient  $r > 0.9$  indicates a strong correlation. A positive sign means that one constituent increases with the increasing of the other one vice versa. Correlation coefficients can be tested with respect to their statistical significance. The “measure of determination”  $100 \times r^2$  (in %) is a better understandable parameter. It cannot be tested statistically. But it explains how many percent of the deviation of all concentration values are attributed to the mutual association: percentage  $> 85\%$  represents a high degree of mutual dependency.

Statistical regression analyses can complete this result only in case of really existing connections: they design a simple or simplified mathematical equation (function) describing the concrete type of connection. The investigator is free to assume a functional type if there are no determined scientific (e. g. physical or chemical) regularities or presumptions. Linear connections as the most simple and best understandable function are preferred and usually set to start.

*Result.* A linear correlation analysis shows

that there are strong linear mutual relationships between the sums of aliphatic and aromatic hydrocarbons in the investigated sample set. The correlation coefficient amounts to  $r = + 0.924$  for 665 involved samples. The “measure of determination”  $r^2 = 85.4\%$  explains a majority of correlatively caused deviation in the cloud of points.

The calculated correlation coefficient  $r$  is statistically significant at a 99% significance level. This means: increasing contaminations caused by aliphatics are generally accompanied by increasing aromatic pollutions and vice versa.

The completing linear regression analysis illustrates type and kind of mutual associations. The resulting regression formula is aromatics =  $0.422 \times$  aliphatics + 48, where aromatics and aliphatics are the concentration values measured in mg/kg.

This function shows that the general content of aromatics in the sample set (the “random sample” built up by all soil samples covering the studied 33 sites in Ogoni) amounts to 42.2% of the aliphatic concentrations added by a constant value of about 50 mg/kg. The constant term has a great significance, first of all, for low and very low concentrations of contaminants. Thus, 0 mg/kg aliphatics will be associated on average with 50 mg/kg aromatics (rounded). 100 mg/kg aliphatics correspond to about 90 mg/kg aromatics; 5000 mg/kg aliphatics are associated with 2160 mg/kg aromatics, etc. All these information is statistically averaged. In any single case – at any concrete place of sampling –, another situation can and will occur. The discussed statistical connection between aliphatics and aromatics is illustrated in Fig. 2. As mentioned, it was possible to select an

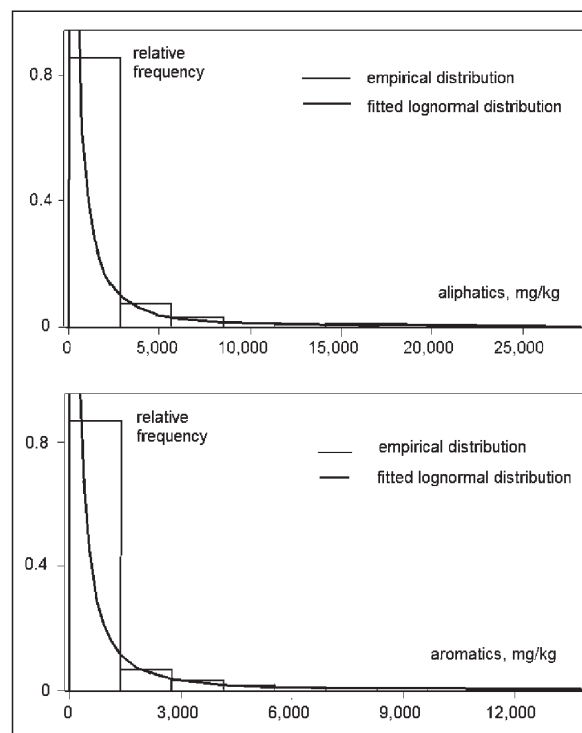


Fig. 1 General frequency distribution of aliphatic and aromatic hydrocarbons

other type of mutual association, e. g. a parabolic or logarithmic one etc. The linear regression is proper to model the relation between aliphatics and aromatics in the Ogoni soil. It is interesting that the selection of a nonlinear quadratic function leads to an analogous result: the curvature of the obtained function is so slight that the visual impression of the curve is the same as at a linear model although the function consists of three instead of two terms at a linear approach. The linear model is easier to interpret and will be preferred here.

### Dependency of Contamination on Sampling Depth

*Topic.* Surface impacts of crude oil into subsoil tend to decrease with increasing depth if the soil is hydraulically permeable. The muddy and clayey subsoil in Ogoni does not belong to well permeable substrates but nevertheless it was expected that the concentrations will decrease with increasing depth. Therefore, it is interesting to prove the dependency of concentrations on the sampling depth.

A second reason to test the dependency of pollution on sampling depth is the fact that the upper subsoil layers have been excavated at many sites, microbiologically remediated and backfilled. Therefore, the original weak stratification of the soil is lost completely. The problem is to determine the effect of backfilling on the depth-related dependency of crude oil concentrations.

*Method.* Bivariate linear correlation analyses between the attributes “concentration” and “sampling depth” were carried out to study possible dependencies. It is the same

Table 5 General statistical properties of main contaminants

	Symbol	Unit	Aliphatics	Aromatics
Number of samples	n	–	665	679
Estimated correct value for the lognormal distribution	$m_l$	mg/kg	3510	1690
Arithmetic mean value (proper for normal distribution only)	m	mg/kg	1360	620
Standard deviation (proper for normal distribution only)	s	mg/kg	$\pm 3010$	$\pm 1360$
Median (position at the concentration axis where a half of data is less and the other half is greater than this value)	–	mg/kg	170	90
Mode (position at the concentration axis where most data occur)	–	mg/kg	10	0
Minimum value	min	mg/kg	0	0
Maximum value	max	mg/kg	28,200	13,800

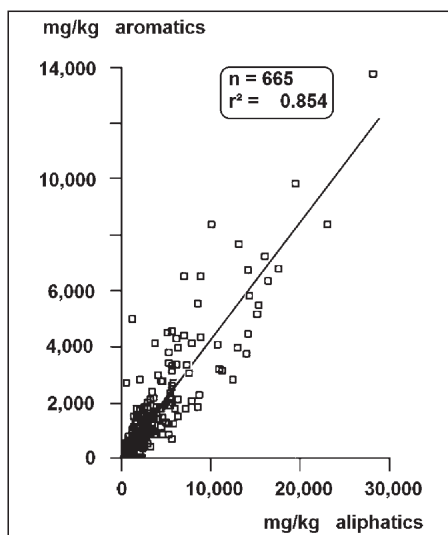


Fig. 2 Statistical dependency between aliphatic and aromatic hydrocarbons

model as applied to study the relation between both parameters and the result will be discussed in the same manner.

Numerous soil samples were taken from the surface and the deepest samples were collected at 300 cm depth (Tab. 1). All soil samples have been taken from (different) sampling intervals. The length of intervals varies between 10 cm and nearly 200 cm. The involved information about sampling depth is therefore only approximated. The number of available samples is differing regarding different pollutants.

**Result.** A significant correlation or a correlation at all between the mean sampling depth and the concentration of aliphatics or aromatics in soil cannot be seen. Only 1.2 % ( $r = +0.110$ ) and 0.5 % ( $r = +0.069$ ) respectively of the deviation of points in the correlogram can be explained by correlation. This does not significantly differ from zero. The result indicates extremely distinct that there are no connections between sampling depth or depth of pollution and the degree of contamination.

So, it seems to be necessary to study the conditions of soil sedimentation. The reasons for the observed missing of any correlation between hydrocarbon concentrations and sampling depths could be twofold. On one hand, the sediments in Holocene deltas themselves show only a weak differentiation or stratification: the retention capacity for contaminants and the degree of pollutants to migrate do not change considerably with changing subsoil depth. Another reason can be the fact that during initial remediation attempts excavated material has been treated by land farming and backfilled only partly remediated. Any possible stratification existing before is gone after such treatment. It also means that mainly backfilled material has been taken for the soil samples at numerous places.

Figure 3 illustrates the missing dependency of contaminating hydrocarbons on the depth of occurrence. It is obvious that dangerous

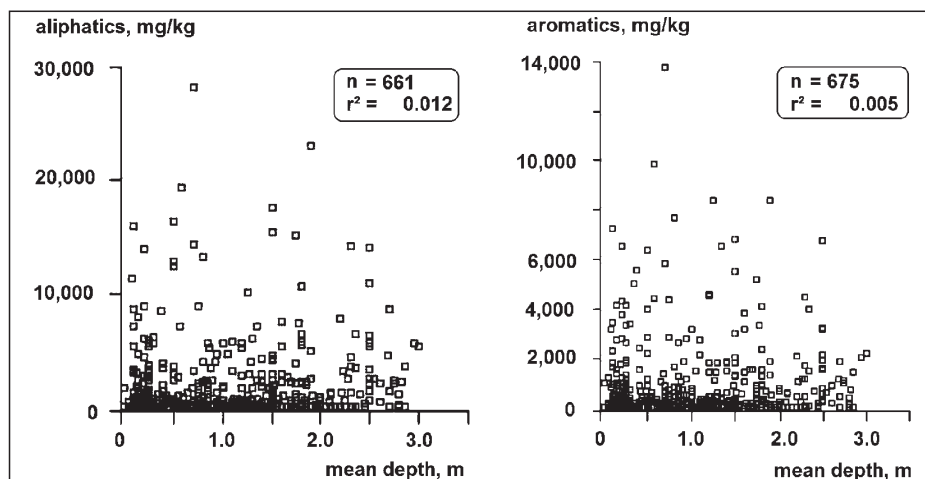


Fig. 3 General connection between polluting hydrocarbons and sampling depth

contaminated soil layers are not limited to near-to-surface layers. They also occur at deeper strata. Furthermore, it cannot be excluded that slightly or not contaminated soil covers strongly polluted horizons which can reach down to the groundwater level in unfavorable cases.

This result confirms the conclusion made in the UNEP report [4] that “in 49 cases, UNEP observed hydrocarbons in soil at depths of at least 5 m. This finding has major implications for the type of remediation required.”

## Conclusions and Recommendations

**Data acquisition.** An economic data acquisition requires to consider a subsequently following computer aided data processing from sampling patterns to laboratory analyses in environmental large scale projects: random sampling is necessary if not specific problems have to be solved; all important contaminants have to be chemically analyzed and information about accuracy and precision of the data is essential; data storage should consider a later possible automated processing from the beginning. It is efficient to determine the kind of attributes to be evaluated in the statistics by preliminary studies.

**Generalizing statistical evaluation.** Environmental risk assessments can be effectively supported and supplemented by a basic mathematical-statistical study if there is a sufficient number of data. Statistical interpretations do not solve punctual problems but inform about the overall characteristics of investigated areas or groups of sites. They must start with the analysis of the distribution function for every involved contaminant to avoid unnoticed errors at the determination of averaging values about the mean degree of pollution. Correlative and regressive studies result in statistical parameters and functions characterizing mutual dependencies inclusively those between contamination and subsoil sampling depth.

A statistical treatment of available data is helpful to prepare decisions about the final

ecological risk assessment according to the general object: to avoid not necessary clean-up measures and expenditures but to revitalize soil and groundwater resources to a necessary extend.

**Detailed study of likely polluted spots and sites.** This essential problem cannot be solved by univariate or bivariate statistical analyses. An application of classifying models is necessary to study such single sampling points or sites in detail (cf. part 3 of this contribution).

**Crude oil contamination in Ogoni.** The total sites studied in Ogoni show that most of the observed sites and single spots are not or only slightly contaminated by crude oil. This can be interpreted as primarily missing pollution or as effect of preceded clean-up measures finished by backfilling of soil. On the other hand, numerous places show unacceptable high pollution in all investigated sampling depths, mainly caused by aromatic hydrocarbons whereas dangerous aliphatic contaminations are comparatively rare. The already initiated processes of soil remediation seem to be incomplete. Nevertheless, monitored microbiological degradation of soil by land farming seems to be a suitable method if carried out correctly.

The authors are grateful to the Nigerian colleagues performing thorough field work at Ogoni to prepare the data basis and also to UNEP for the systematic survey of more or less polluted sites in the studied region.

## References

- [1] Shell: Environmental Quality Standards – Soil and Groundwater. EP 95-0385. HSE Manual, The Hague/The Netherlands, 1995, 33 pp.
- [2] Thiergärtner, H.; Holtzmann, K.: Mathematically aided risk assessment of crude oil contamination in Ogoni, Nigeria. Part 1. Geo-ecological characterization of the pollution. OIL GAS European Magazine 37, III/2011, pp. 143–147.
- [3] UNEP: United Nations Environment Program. Ogoniland Environmental Assessment. Geneva/Switzerland, 2011 [www.unep.org/documents/...].
- [4] UNEP: Environmental assessment of Ogoniland. Abuja/ Nigeria, 2011, 246 pp., [www.unep.org/Nigeria/01\_fwd\_es\_ch01\_UNEP\_OEA.pdf to www.unep.org/Nigeria/08\_app\_UNEP\_OEA.pdf]