International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 21(2), 2021, pp 1-18

Environmental Statistics – A Pet Subject of Sinha Brothers*

T. J. Rao

Retired Professor, Indian Statistical Institute, Calcutta, India

Email: tjrao7@gmail.com

[Received April 27, 2021; Accepted September 25, 2021]

Abstract

In this paper, we shall familiarise the readers with certain concepts related to environmental statistics – covering both environmental sampling and inference - a pet subject of both Sinhas. We shall also discuss certain hazards and risks that are associated with environment. This article is aimed at research scholars, practitioners and others interested in environment such as NGOs and policy makers and hence no advanced knowledge of sampling is assumed.

Keywords: Environmental sampling, Basic sampling techniques, Adaptive and network sampling, Sample size, Hotspots, Hazards and risks, Data banks.

AMS Subject Classification: 62D05, 62P12.

0. Prologue

Bimal Kumar Sinha (BMS) and Bikas Kumar Sinha (BKS), the twin brothers excelled in academics, choosing the subject of Statistics for their career. Prasanta Chandra Mahalanobis, founder of the Indian Statistical Institute (ISI) is well known for his innovative contributions in Sample Surveys as well as Multivariate Analysis. Widely applied methodologies popularly known as Mahalanobis D^2 Statistic and Inter Penetrating Network of Subsamples (IPNS) among several other PCM's contributions are Applied and Survey Statisticians' favourite tools of

---

*Paper based on lectures given at the International Workshop on *Statistical Methodology for the study of Environmental Data and Climate Changes* at the C. R. Rao AIMSCS, Hyderabad in March 2014.

Data Analysis. Thus after graduating from Calcutta University, during their tenure at the ISI, it is of no surprise that the twins got inspired by the founder's work. While BMS chose Multivariate Analysis, BKS concentrated on Sample Surveys.

Hailing from, the then, East Bengal, environmentally rich and ecologically diverse background, it is but natural that both of them turned their attention to 'Environment' along with their specialised areas of Statistics.

Bimal Sinha's (BMS) research has been funded by the US Environmental Protection Agency for about two decades and his research contributions have been recognized through a *Distinguished Achievement Award* from the Environmental Statistics Section of the American Statistical Association. Some of his research work on environmental risk analysis in Fonseca *et al.* (2015), rigorous statistical methodology for Physiologically-based Pharmacokinetics (PBPK) in Klein *et al.* (2012), effects of Formaldehyde exposure on the health risk in Bateson *et al.* (2012), pre-plant soil fumigation in Knapp *et al.* (2009) are well known and often cited.

BMS published a series of papers during late nineties with EPA statisticians on various aspects which include evaluating cleanup standards, parametric and nonparametric. Furthermore, a rigorous statistical analysis of Hillsdale Lake (situated in Kansas state) data was performed by BMS joint with his doctoral student Dr. Xiaoming Li and EPA Chief Statistician Dr. Barry Nussbaum in 2000, and this paper published in Calcutta Statistical Association Bulletin (2000) was awarded Bose-Nandi prize for being the best Applications paper in that year.

Bikas Sinha (BKS) was appointed as 'Expert on Mission' for United Nations (UN) Statistics Training Programme in 1991 and this led to his serving as a consultant to the United States Environmental Protection Agency (USEPA) in 1993. His research contributions appeared in *Encyclopaedia of Environmetrics* (2002), *Journal of Environmetrics* (2003) and *Scandinavian Journal of Forest Research* (2004) to quote just a few. Besides collaborating with around 110 co-authors worldwide on a wide range of Statistical problems, he maintained his interest in environmental projects such as estimation of unknown population size, wild elephant counts in Assam sanctuaries, footfalls at exhibitions and temples, social networks and cultural diversity, He was also appointed a Member of the apex body, National Statistical Commission, Government of India. (2006-2009). His expertise involved in social and environment statistics in such areas as population, health, education, labour and employment or environment. He was also involved in drafting the 'Collection of Statistics Act, 2008' in defining 'core statistics' and classifying the same into various categories and subcategories. The act empowers collection of statistics on various aspects including 'environment' through a sample survey in which BKS excelled.

BKS was awarded the P.C. Mahalanobis Gold-rimmed Silver Medal by Sadharan Brahma Samaj, Kolkata in 1980 for contributions in Survey Sampling and 'Centenary Medal of Excellence' by the School of Tropical Medicine, Kolkata in 2004.

1. Historical references to Environmental Sampling in India

It is interesting to note that the great Indian epic Mahabharata mentions that during a long ride in a dense and well preserved forest, king Nala (in disguise) and king Rituparna break their journey under a tree for the night and Rituparna trying to flaunt his skill in environmental sampling, estimates the number of fruits on the tree as 2095. Nala makes a complete count and is duly amazed by the correctness. Rituparna accepts his due: "*I, of dice possess the science, and in numbers thus am skilled*."

In Kautilya's *Arthashastra*, attributed to 321 B.C., it is mentioned that the Mauryas "*looked at the forests as a resource*". Further, *Arthashastra* unambiguously specifies the responsibilities of officials such as the *Protector of the Elephant Forests*:

"On the border of the forest, he should establish a forest for elephants guarded by foresters. The Superintendent should with the help of guards...protect the elephants whether along on the mountain, along a river, along lakes or in marshy tracts...They should kill anyone slaying an elephant".

Next, coming to the Moghul Period, we note that Babur offers his description of fauna of India in a very systematic style. After giving the features of India's physical geography, he proceeds to describe first the mammals, then birds, and, finally, aquatic animals.

Jahangir ordered his artists to portray animals and birds as well as to prepare accurate paintings of flowers providing all the botanical details necessary for identifying the family of the flower.

In *Ain-i-Akbari*, Abul Fazl (circa 1590 A.D.) describes the Flora and Fauna in detail. The Moghul Emperors and nobles had another interest that indirectly brought them closer to ecology. Laying of gardens whether resting places and parks or flower gardens (*gulistan*), (*bostan*) and orchards was a favourite pastime of the kings, princes, princesses, and nobles.

During his reign in the 17th century, Maratha Emperor Shivaji showed his compassion for environment and seems to have issued a 'Letter of Order' (*Ajnapatra*) for protecting trees in the forest.

Turning to a recent British Period, Francis Buchanan's Survey of Eastern India, 1807 describes two surveys he conducted, the first of Mysore in 1800 and the second of Bengal in 1807-14. In 1804, he was in charge of the newly founded 'Institution for promoting the Natural History of India' at Barrackpore near Calcutta. He made a comprehensive survey of Bengal and reported on topography,……., natural productions (particularly fisheries, forests, mines, and quarries), agriculture,……. His research includes an important work on Indian fish species, entitled *An account of the fishes found in the river Ganges and its branches* (1822), which describes over 100 species not formerly recognized.

Coming to the recent past, we remark that of the several studies and projects undertaken by the ISI, the author along with the other team members had the privilege of working with BKS on many such topics relating to environment. Some of the important ones include survey methodology for estimation of the number of wild elephants in the states of Arunachal Pradesh and Assam, estimation of tiger count in the Sunderbans of West Bengal, footfall at annual events such as book exhibitions (*boi mela* in Calcutta Maidan) and festivals at temples and in general, statistical inference for the estimation of unknown size under various settings, social networks. In addition to working on these environment statistical projects , the author and BKS collaborated in tracing the evolution of Statistics in India (Ghosh *et al.*,1999) as well as a brief history of Statistics and its development in the Indian subcontinent (Rao and Sinha, 2011) .

Fast-forward to the present day, and we now have the Environment (Protection) Act, 1986 in India which oversees 'the protection and improvement of human environment and the prevention of hazards to human beings, other living creatures, plants and properties'. Similarly, we have the Bangladesh Environment Conservation Act (BECA), 1995 which proposes a set of laws to 'provide for conservation of the environment, improvement of environmental standards and control and mitigation of environmental pollution'.

2. Types of data and Cross Examination: *Look before you jump to computations*

Scientists are generally concerned with different types of data. Some of these refer to geography, geology etc. Most of the data relates to a point of time and sometimes to a period of time. The first type is more or less constant such as the topography of districts, rivers, locations of mineral reserves and so on. Once the data is collected, for example, by Survey of India or geological expeditions there is no need to collect it for the next few years. To collect data over a period of time, one needs to collect data at the initial point of time and at the relevant end point. Thus one needs to monitor data collection at a point of time through a well defined methodology. However, there are certain types of data which can be obtained only through a well designed experiment such as finding out which combination of N,P and K would give a better yield of a crop. In this paper, we shall look at the methodology for collecting data mostly at a point of time.

When a scientist or a statistician is collecting data with a specified objective in mind we call it primary data, while if the aim is to utilize already existing data (collected by some one else, such as Population Census of the Office of the Registrar General of India or Large Scale Sample Surveys such as those conducted by National Statistical Office (NSO) of the Indian Ministry of Statistics and Programme Implementation (MoSPI) or Population Census conducted by Bangladesh Bureau of Statistics (BBS), we call it secondary data. While in both cases, it is very important to *cross examine* data before drawing inferences based on them, as emphasized by R. A. Fisher, the Father of Modern Statistics, P.C. Mahalanobis, the Founder of the Indian Statistical Institute and C.R. Rao, the Doyen of Statistics, one has to scrutinize more carefully the secondary data collected by someone else before making use of it.

While analyzing Bengal anthropometric field data of 1945, C.R. Rao emphasizes that

"...Nothing is more frustrating to the investigator than to discover that the observations collected at a considerable expense of money and energy are

worthless because of obvious inconsistencies or failure to furnish complete details…"

In his book *Statistics and Truth* (Rao, 1989), C.R. Rao discusses strategies for cross examination of data (CED) including detection of faking of data in environmental statistics among others. CED is regarded as an important guide to many scientists working on secondary (or even primary) data. We shall illustrate this by a couple of examples:

***Example* 1:**

A question is set as follows: An environmentalist recorded maximum temperatures from 3-9 May, 2020 in Kolkata as 33, 36, 35, 33, 35, 36 and 38 while in Kota as: 28, 25, 25, 27, 25, 26 and 30.

Which collection of data is **more** variable, assuming both are measured in Deg. Celsius?

***Example* 2:**

Heights of four boys of class ten in a school are: 5ft.4in., 5ft.6in. 5ft.4in. and 4ft.9in., what is their mean?

In the first example, before jumping to a quick conclusion that the temperatures in the second list [Kota] are less variable than in the first list [Kolkata], one should question the correctness of the data for Kota which are *not at all likely to be* as given here, the time specified being a mid-summer in the Indian State of Rajasthan, a hotter region. Here, the data is wrongly recorded. In fact, instead of the maximum temperatures, the environmentalist has recorded the minimum temperatures for Kota, possibly unmindfully!

For the second example, a good mathematician will add all the four values and divide by four, while a good statistician will notice that the last value is an outlier and delete and calculate the average of the first three and call it a modified mean. On the other hand, an environmental specialist would contact the family of the fourth child and find out whether there is any environmental effect that caused the stunted growth, such as a pesticide factory nearby or high tension power lines or low nutrition etc. before calculating the mean.

Example 3:

(*This example is a variation of an actual problem faced by C R Rao during the analysis of data from Bengal Anthropometric Survey: 1945. Since the 1945 observations are unfamiliar to many readers, we have chosen different numbers to drive home the main point.)*

9.0	8.4	9.5	9.5	8.5	9.4
9.3	9.2	9.3	9.3	9.0	9.0
9.1	9.3	9.4	9.4	9.1	9.3
8.5	9.4	9.2	9.5	9.4	9.5
9.4	8.5	9.0	9.2	9.1	9.2

Closely examine the data above and calculate the average.

Immediate reaction would be to add up the entries which comes to 274.9 giving an average of 9.1633.

In this example, the 'point' is not a decimal. It is like a comma. On a closer examination, after the 'point' we do not find a 6, 7, 8 or 9. They are all noted as modulo (6) by the statistician. In fact, this data refers to the number of overs bowled by a player in a one- day cricket match. Hence they should be totalled 'modulo 6' and then average taken.

Primarily, inference is based on descriptive statistics computed from the data such as measures of central tendency, namely mean, median , mode or measures of dispersion like standard deviation, range etc. Inference based on analytical statistics answers questions by testing hypotheses as evidenced by the data.

3. Environmental Data and Data Analysis

Consider the following types of data:

a) Usage of plastic bags

b) Printing of train/flight tickets.

c) Printing reprints of published papers to read later

d) Leave the lights on in the office during absence from the room.

If a person satisfies any of the above criteria, he or she is considered/identified as *environment unfriendly*.

In the above examples, the parameters of interest are the proportions of individuals engaged in such anti-environmental activities. As is well-known, there are standard sampling procedures to estimate such proportions to any sufficient degree of accuracy by choosing a suitable sample size. Vide, Hedayat and Sinha (1991). However, in environmental sampling, extra care should be taken in the selection of sampling designs before carrying out data analysis.

4. Biased Selection of Samples

An environmentalist who visits temples often, has taken water samples from Godavari river near Kopergaon, Baasara and Rajahmundry [South India] and gets it analysed for pollution levels. Based on the estimates from this sample, he writes a paper on the *Pollution of River Godavari*. Comment on this approach and suggest a proper methodology.

Here, choice of purposive sample locations is *not* correct. The first one is nearer to the birth place of the river and the last one is nearer to the estuaries. One could visualise a similar situation with reference to *River Padma. In case of spatial sampling, uniformity of the sites is to be strictly adhered to.*

5. Certain other sampling designs

Standard sampling designs are discussed at length in text books. We may refer to Hedayat and Sinha (1991), for example. We will not dwell on such sampling methodologies any further. Instead, we will introduce some sampling aspects specific to environmental assessment issues.

Adaptive and network sampling designs

In certain situations, it is very difficult to get a sampling frame. In that case one has to build up a suitable frame. For example let us say that we are interested in factories on a riverside which are polluting the river. We can add units one by one by ascertaining from the factory workers who else in that area is indulged in polluting. Thus we build up a network which forms the basis for an adaptive sampling. Similarly, when dealing with animal abundance, the sampler keeps on adding units to the sample whenever he comes across an abundance. On the other hand, if we are interested in the prevalence of a rare characteristic in a population of households, the adult members are asked to report not only if they possess that characteristic, but whether their siblings also possess this and thus a network of units is obtained so as to utilise a network sampling technique. Vide Thompson and Seber (1996).

Capture-recapture techniques

A technique that is adopted to estimate the unknown population size itself is known as capture- recapture technique. Suppose that we wish to estimate the number N of fish in a pond. A simple version of the technique, under certain assumptions is: Capture a first sample of M fish, mark and release them. After a while, take a second sample of size m and let r be the number recaptured. An unbiased estimate of N is given by

$$\hat{N} = \{(M + 1) (m + 1) / (r + 1) \} – 1.$$

Some of the other practical techniques like *Line Intercept Sampling*, *Composite Sampling* among others are discussed in Thompson (2002) .

There is a huge literature along this line of research – both theory and application. BKS and his collaborators have worked extensively in this area. We refer to De and Sengupta (2003), Goodman (1953), Gupta (1967), Hossain (1995), Samuel (1968), Sengupta and De (1997), Sinha and Bose (1985) and Sinha and Sinha (1975).

Weighted distributions

Noticing the concept of weighted distributions first mentioned by Fisher, C.R.Rao(1965) formulated it in general terms for statistical data on modelling, when the conventional distributional assumptions do not hold especially in ecological studies. This concept is found to be useful in certain sample survey situations, ecology, medicine, reliability (see Patil and C.R.Rao (1978) and Patil (2002)).

Sample Size Determination

***Example* 4:**

Based on a Simple Random Sample selected WithOut Replacement from a population of N = 491 lakes in a region, it is desired to estimate the average amount of garbage dumped into these lakes. The scientist prefers to have the coefficient of variation (c.v.), g of the estimate not more than 5%. From a previous study, it is known that the region has a c.v. of 0.2771 for this characteristic. Find the sample size required.

Solution. Using the formula, we get $n = 1 / \{ (1/491) + ((490/491) ((0.05)^2)/ (0.2771)^2 \} = 29$. Formula for large N, gives the required size as $n = 31$. See Murthy (1967) and Rao (2014) where the formula

$$n = 1 / \{ (1/ N) + ((N - 1)/ N)(g/ \eta)^2 \}$$

where g is the pre fixed r.s.e. and η is the knowledge of c.v. of the population and other details are discussed.

Using a probability requirement :

Sometimes, it is desired to obtain the sample size when the relative error (r.e.) in the estimated population mean using a design, say SRSWR, is to be controlled at a given level, say r. Further, this r.e. can be allowed to exceed with a chance of α%. Here r.e. $= (\bar{y} - \bar{Y}) / \bar{Y}$ and Prob. $\{- r \leq ((\bar{y} - \bar{Y}) / \bar{Y}) \leq + r \} = 1- \alpha$.

Assuming normality of y- values which gives $\bar{y} \sim N (\bar{Y}, \sigma/\sqrt{n}))$ write

$$\text{Prob. } \{- z \leq (\bar{y} - \bar{Y})/ ((\sigma/\sqrt{n})\} \leq + z \} = 1- \alpha.$$

Comparing the two Prob. Statements we get, $r \bar{Y} / (\sigma/\sqrt{n})\} = z$ giving $\sqrt{n} = (z/r) (\sigma/ \bar{Y})$ from which n could be obtained.

Here we should have an idea of population C.V. $= (\sigma/ \bar{Y})$ from past data, pilot survey or related auxiliary information as in the *previous* example. For details we refer to Murthy (1967).

6. Detection of Hot spots

Square, rectangular or triangular grids are used and samples are collected at the grid nodes. Of these triangular grids are found to be more efficient. It is important to determine the grid spacing and refer to the nomograms which are available. From these nomograms, size of hot spot, probability of 'no hit', etc. can be found.

Also the chance that a hot spot exists when there is ‘no hit’ can also be calculated by using a simple application of Bayes Theorem. For further details, we refer to Gilbert (1987).

7. Assessing hazards - a few illustrations

We shall now look at a couple of examples to illustrate hazard assessment.

This assessment with respect to glacial lakes in Himalayas is an important problem on which certain attempts are made in the current literature. However, in one such attempt, the models used for the probability of such an occurrence using regression parameters of a logistic fit are meant for Canadian data which may not suit Indian conditions. Thus one has to answer questions such as, whether we have enough data, whether the assumptions of the model are valid, whether the model itself is relevant etc. Recent research includes identification of hazard and risk for Glacial Lakes in the Nepal Himalayas using satellite imagery which was presented in Rounce *et al.* (2017). Furthermore, hazards from Himalayan Glacier Lake Outburst Floods (GLOF) were recently described in Veh e*t al*. (2020). The very recent GLOF of Nandadevi Glacier near Chamoli in the Indian State of Uttarakhand on 7 February 2021 is a case in point. Whether it is caused by climate change or developmental construction in a fragile ecosystem is being investigated.

Arsenic (As) contamination especially in the Ganges Bengal Delta Plain (BDP) which became a health hazard for more than 50 million villagers in the state of West Bengal itself, has received much attention by environmental researchers (see, for example, Singh (2006), Bhattacharya *et al.* (2001), Chakraborti *et al.* (2003) among several others). There are several hot spots (As > 0.01 mg/l as per WHO standards and As > 0.05 mg/l as per BIS) in the middle and lower ranges of the river Ganges. Though arsenic contamination due to natural geo-chemical processes such as volcanic eruptions or suspended particles in air circulation can not be controlled, it is the industrial pollution of rivers, underground waste disposals that are causes of concern for the environmentalists. Further, another arsenic pollutant, namely, Chromated Copper Arsenate (CCA) in wood processing is still being used in several countries, even though United States EPA, Canadian and European agencies have restricted this.

Referring to the arsenic pollution in the central and lower regions of the river Ganges, we note that not many research studies are available from the statistical point of view. Though several guide lines are prescribed for collecting samples, it is usually the “grab samples” which are taken for laboratory analyses. Once the major locations of hot spots are known, one could use auto sampler devises for collecting water samples at, say, systematic sampling time intervals. Systematic sampling can be used for selecting the dates of a month as well, starting with a random start date. One of the earlier statistical studies in the eighties, w.r.t. arsenic pollution in the river Ganges was taken by by Somesh Dasgupta of ISI, Kolkata. Purkait *et al.* (2008) used an artificial neural network (ANN) model as multilayer perceptron (MLP) architecture to estimate arsenic concentration in ground water of Malda district. In continuation, Sengupta *et al.* (2010) analysed data based on 700 drinking water samples from tube wells in the hot spots of deltaic alluvial plains of Bengal delta in Malda district. They prepared As zonation map with 6 groups. Statistical tests for isotropy (circular uniformity/randomness) against preferred direction of contamination of arsenic as well as Change Point (CP) test to detect angle at which contamination changes, were used. However, for selecting sample points in a polluted river care has to be taken so that proper statistical analyses can be made. For monitoring pollution along a river, a mathematical formulation of the problem of optimally selecting the sample points was given by Alvarez-Va'zquez *et al.* (2006). They also provided an efficient algorithmic solution.

Bangladesh Bureau of Statistics conducted a Bangladesh National Drinking Water Quality Survey (BNDWQS) in 2009 in which 15,000 random clusters were sampled in the country and samples were collected from 300,000 hhs. The report noted that 13.4% of water samples are contaminated by Arsenic. Hasan *et al.* (2019) compiled recent data from various sources on trace metal pollution in Bangladesh. They also discuss its impact on public health.

Next, we shall refer to the hazards of natural floods. In 1926, a catastrophic flood occurred in the Brahmani river in Orissa and several low-lying areas got flooded. It was decided by a group of engineers that the river bed had risen and hence the height of embankments should be raised by several feet. The matter was referred to P.C. Mahalanobis of the Indian Statistical Institute who did several statistical analyses. Mahalanobis found a significant correlation between the rainfall in the catchment area and height of the river flood in the delta area. His report of 1930 to the government of Orissa led to the construction of the multi-purpose project Hirakud dam, thus avoiding the raising of the embankments and establishment of Damodar Valley Corporation. This early work in Operations Research is perhaps very much applicable in similar situations in any country.

In 2018, heavy monsoon rainfall in Kerala resulted in the onset of flooding towards the end of July. A severe spell of rainfall was experienced at several districts for two consecutive days on the 8th and 9th of August. This led to further uncontrollable flooding in the State. Due to unprecedented monsoon and heavy rainfalls in catchment areas, water reached the Full Reservoir Level in many reservoirs. This necessitated the release of water from several dams due to heavy rainfall in their catchments. There has been a debate about the flooding due to simultaneous release of dam waters. This could perhaps be solved as a traveling salesman problem (TSP). Analysing the past data on rainfall at the reservoirs, their thresholds and danger levels and other relevant data, it might be possible to obtain an optimum route of opening the dams rather than simultaneous release at the same time.

Biodiversity indices such as Shannon's diversity, Simpson's diversity etc. are considered in the literature on the subject , but in the context of India or Bangladesh, not much information is available due to the complex nature of the problems.

Before concluding this section, we mention now an important work carried out by Sinha brothers. This refers to cleaning of hazardous nuclear waste sites. While BMS was a faculty at UMBC and attached to US EPA as a statistics consultant, he had inducted BKS as a summer-visitor at UMBC and they collaborated on the following problem:

Hazardous Nuclear Waste Sites need cleaning at regular intervals as otherwise nuclear waste creates an alarming threat to the surroundings. When the pollution level exceeds a critical value, waste cleaning is initiated. Once the cleaning operation starts, the pollution level starts declining and when it reaches an acceptable tolerance level, the cleaning operation is stopped temporarily. In order that this costly and dangerous exercise of cleaning a hazardous nuclear waste site is fruitful, one has to study and monitor the pollution level before and after cleaning. Therefore, in principle, one has to collect nuclear pollution data from randomly selected spots in the site both before and after cleaning. As regards development of relevant statistical theory, Sinha brothers embarked on the problem of joint modelling of bivariate data, say X (before cleaning) and Y (after cleaning). There is an inherent restriction on (X, Y) in the sense of ($0< Y < X <$ infinity) apart from being correlated. This modelling of bivariate non-negative restricted random variables with/without specified marginals is itself an interesting exercise.

They developed the model and thereafter, suggested a testing procedure for the hypothesis

Ho: $\bar{Y} = f \ \bar{X}$, where '*f*' refers to a specified fraction. For this interesting and important study we refer to Sinha and Sinha (1995). Apart from this, they have another joint work (Sinha and Sinha (2002)) on Multi-Criteria Decision Making (MCDM) methods with applications to environment.

The famous Superfund Project at USEPA involved allocation of billions of dollars to the 50 US states to carry out extremely expensive and most complicated cleanup jobs at the environmentally hazardous waste sites prevailing in USA. This required ranking of the states in order of their exposing the public to health hazards. The ranks were based on suitably combining four indices comprising air pollution, groundwater pollution, surface water pollution and land pollution. The novel method of Multiple Criteria Decision Making (MCDM) was used to accomplish this job and BMS (and later BKS) wrote a few papers on this topic (2003, 2004).

In the context of environmental risk assessment, laboratory data based on animal studies are usually used for model building and risk assessment (absolute risk, additional risk, extra risk and benchmark dose). Resulting results are subsequently extrapolated to humans. Physiologically based Pharmaco-kinetic models are often used in this scenario. BMS along with a few EPA statisticians contributed to this important and very useful area (2011, 2012).

Dose-response modeling for continuous responses based on multistage Weibull models is frequently used in environmental applications. While it is mostly assumed that the variability remains constant with increase in dose, there are many practical EPA studies where this does not hold and one needs to address this concern. BMS along with two EPA statisticians (John Fox and Leonid Kopylev wrote a few theoretical papers which were published in international journals ([7], [17]).

8. Epilogue

In the Indian context, some of the data inventories available are: Toxic Release Inventory (TRI), inventory of hazardous chemicals imported in India, national wetland inventory and assessment (NWIA), national forest inventory in India, Inventory of aerosol and sulphur dioxide emissions from India etc.. It may also be noted that the Environment Statistics Section of the United Nations Statistics Division (UNSD) collects environmental data and environmental indicators derived from these data are published in ENVSTATS. Apart from UN Environmental Program Data Portal, World Bank, Earth Science Information Network are some other sources of data. Ministry of Statistics and Programme Implementation of NSO's recent publication ENVISTATS INDIA, 2019 is available in 2 volumes. Similarly for Bangladesh, World Bank, UNDP, ADB and other organizations publish data. Moving from the earlier Millennium Development Goals (MDG) to Sustainable Development Goals (SDG), all countries now have an agenda before them to reach certain targets. SDG 13 refers to climate change and its impacts, SDG 14 deals with conservation and sustainability of ocean and forest resources and other terrestrial ecosystems. It may be noted that MoSPI in their Progress Report of 2020 on SDG National Indicator Framework gives the details of achievements during 2019-20 with respect to Goals 13, 14 and 15. The country has also reduced its voluntary contribution of emission intensity by 21%. Similar data will be available for Bangladesh in the reports. The United Nations Convention to Combat Desertification (UNCCD) COP 14 was held in India in September 2019 and the parties to the convention agreed on the agenda for the next two years and beyond to lead to a sustainable development path.

While there are efforts taking place by international agencies, let us hope that Sinha brothers continue their interest and involvement on posing and solving the problems relating to environment using their excellent expertise in Statistics. We congratulate and wish them a very Happy Seventy Fifth Birthday Celebration.

References

[1] Alvarez-Va'zquez, L. J., Martinez,A., Va'zquez-Mendez, M. E. and Vilar, M. A. (2006). Optimum location of sample points for river pollution control, Maths. and Computers in Simulation, 71, 149-160.

[2] Bateson, T., Glenn, B., Bayliss, D., Xu, D. and Sinha, Bimal, K. (2012). Epidemiology, Meta analysis of effects of Formaldehyde exposure on the risk of spontaneous abortion, 23, P- 373.

[3] Bhattacharya, P., Jacks, G., Jana, J., Sracek, A., Gustafsson, J. P. and Chatterjee, D. (2001). Geochemistry of the Holocene alluvial sediments of Bengal Delta Plain from West Bengal, India: implications on arsenic contamination in groundwater. In Proceedings of National Seminar on Groundwater Arsenic Contamination in the Bengal Delta Plains of Bangladesh (eds. Jacks, G., Bhattacharya, P. And Khan, A. A.), University of Dhaka, Bangladesh, TRITA-AMI Report, 2001, 3084, pp. 21–40.

[4] Chakraborti, D., Mukherjee, S. C., Pati, S., Sengupta, M. K., Rahman, M. M., Chowdhury, U. K., Lodh, D.,Chanda, C. R. ,Chakraborti, A. K. and Basu, G. K. (2003). Arsenic Groundwater Contamination in Middle Ganga Plain, Bihar, India: A Future Danger?, Environmental Health Perspectives 111, 1194-1201.

[5] Das Gupta, S., Unpublished Tech. Report, Indian Statistical Institute.

[6] De, M. and Sengupta, S. (2003): Bernoulli Sequential Estimation of the size of a finite population, Sequential Analysis, 22, 95 -106.

[7] Fonseca, M., Mexia, J.T., Sinha, B.K. and Zymslony, R. (2015) .Likelihood ratio tests in linear models with linear inequality restrictions on regression coefficients, Revstat- Statistical Journal, 13 ,103-118.

[8] Fox, J., Frazier, E. and Sinha, Bimal K. (2011). Dose-Response Modeling for continuos responses: Alternative variance models, Int. Jour. Stat. Sci., 18, 189-206.

[9] Ghosh, J. K., Maiti, P., Rao, T. J. and Sinha, B. K. (1999). Evolution of Statistics in India, Int. Stat. Rev., 67,13-34.

[10] Gilbert, R. O. (1987). Statistical Methods for Environmental Pollution Monitoring, Chapter 10, Wiley, N.Y.

[11] Goodman, L. A. (1953). Sequential sampling tagging for population size problems. Ann. Math. Statist., 24, 56 - 69.

[12] Gupta, M. K. (1967). Unbiased estimate of 1/p, Ann. Inst. Stat. Math., 19, 413-416.

[13] Hasan, Md. K. Shariar, A. and Jim, K. U. (2019). Water pollution in Bangladesh and its impact on public health. Heliyon. 5, e02145.

[14] Hehayat, A. S. and Sinha, B. K. (1991). Finite Population Sampling, Wiley, N.Y.

[15] Hossain, M. F. (1995): Unknown population size estimation: an urn model approach, Journal of Statistical Studies, 15, 89-94.

[16] Klein M. D, Sinha, Bimal K, Subramaniam R. P. (2011). Statistical Inferences from formaldehyde DNA-protine cross-link data: improving methods for characterization of uncertainty. Jour. Biopharm Stat., 21, 42-55.

[17] Klein, M., Neerchal, N. K., Sinha, B. K., Chiu, W. A. and White, P. (2012). Statistical inference from serially correlated Methelyne Chloride data, Sankhya, B, 74, 211-237.

[18] Kopylev, L. and Sinha Bimal K. (2011). On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary. Sankhya, B,73, 20-41.

[19] Knapp. G., Sinha, B. K. and Xu, D. (2009). Extracting within experiment precision of horticultural experiments useful for Meta-Analysis, Jour. Appl. Horticulture, 11, 10-17.

[20] Maitra, R., Ross, P. N., Sinha, Bimal K., Lee, J. and Herczeg, S. (2003). On some aspects of data integration techniques with applications. East-West Journal of Mathematics, Computational Mathematics and Modelling: CMM:241–256.

[21] Montip, T., Satinee, L. and Sinha, Bimal K. (2005). On a comparison of four estimates of a common mean by multiple criteria decision making method. Jour. Stat. Res., 39, 17-28.

[22] Murthy, M. N. (1967). Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

[23] Nussbaum, B., Li, X. and Sinha, Bimal K. (2000). A statistical analysis of Hillsdale lake data. Cal. Stat. Assoc. Bull., 50, 293–305.

[24] O'Brien, R., Smith, W. and Warren, J. (1994). Some aspects of simultaneous inference. Comm. Stat., Theory & Methods, 23, 59–87.

[25] O'Brien, R., Sinha, Bimal K and Smith, W. P. (1991). A statistical procedure to evaluate cleanup standards", Jour. Chemometrics, 5, 249–261.

[26] Odom, B. and Sinha, Bimal K. (1995). Asymptotic semiparametric procedures to evaluate cleanup standards. Jour. Appl. Stat. Sci., 2, 89–102.

[27] Patil, G. P. (2002). Weighted distributions, Encyclopedia of Environmetrics, 4, 2369–2377.

[28] Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size biased sampling with applications to wildlife populations and human families, Biometrics, 34, 179- 189.

[29] Purkait, B., Kadam, S. S. and Das, S. K. (2008). Application of artificial neural network model to study arsenic contamination in groundwater of Malda District, Eastern India. Journal of Environmental Informatics, 12, 140-149,

[30] Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment,. Sankhya, A, 27, 311-324.

[31] Rao, C. R. (1989). Statistics and Truth. Ramanujan Memorial Lectures, Council of Scientific and Industrial Research, India

[32] Rao, T. J. (2013). Environmental Sampling techniques- A Brief Introduction, Report 2013-14, C.R Rao AIMSCS, Hyderabad.

[33] Rao, T. J. and Sinha, B. K. (2011). A brief history of Statistics and development in the Indian Subcontinent, Int.Jour.Stat.Sci., (Special issue in honour of P.C. Mahalanobis, 11, 1-15.

[34] Rounce, D. R., Watson, C. S. and McKinney, D. C. (2017). Remote Sensing 9, 654.

[35] Samuel, E. (1968). Sequential maximum likelihood estimation of the size of a population. Ann. Math. Statist., 39, 1057-1068.

[36] Santinee. L., Montip, T. and Sinha, Bimal K. (2004). A statistical approach to combining environmental indices with an application to air pollution data from Bangkok, Thailand. Pakistan J. Stat. 20, 245- 261.

[37] Sengupta, S. and De, M. (1997): On the estimation of size of a finite population, Sankhya, Series B, 9, 66-75

[38] Singh, A. K. (2006). Chemistry of arsenic in ground water of Ganges-Brahmaputra river basin. Current Science, 91, 599-606.

[39] Sinha, Bimal K., Li, X. and Nusbaum, B. (2003). Air pollution statistics and environmental health perspectives . Sustainable environments -A statistical analysis. In Ghosh, A. K., Ghosh, J. K. and Mukhopadhyay, B.(eds.). Oxford University Press, 63–86.

[40] Sinha, Bikas, K. and Bose, A. (1985). Unbiased sequential estimation of 1/p: Settlement of a conjecture. Ann. Inst. Statist. Math., 37, 455-460.

[41] Sinha, B. K. and Sinha, B. K. (1975). Some problems in unbiased sequential binomial estimation. Ann. Inst. Statist. Math., 27, 245-258.

[42] Sinha, B. K. and Sinha, B. K. (1995). An application of bivariate exponential models and related inference., Jour. Stat. Plann. and Inf. , 44 , 181--191.

[43] Sinha, B. K. and Sinha, B. K. (2002). On some aspects of MCDM with environmental applications. In honour of the 65th birthday of Professor Mir Masoom Ali, J. Stat. Stud. 2002, suppl., 139--144.

[44] Thompson, S. K. (2002). Sampling, Wiley, N.Y.

[45] Thompson, S. And Seber, G. A. F. (1996). Adaptive Sampling, Wiley, N.Y.

[46] Veh, G., Korup, O. and Watz, A. (2020). Hazards from Himalayan glacier lake outburst floods. Proc. Nat. Acad. Sci., 117, 907-912.