Observation Strategy

K. Shimizu

(Science University of Tokyo)

Observation Strategy

Kunio Shimizu
Department of Applied Mathematics
Faculty of Science
Science University of Tokyo
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162

1. Statistician's point of view

Many statistical books discuss the design of experiments, considerations effecting the method of making, or collecting, the observations to be analyzed. However, it is true that there are many cases where experimental conditions have been set up by physical, mechanical and other reasons in practice. Even under such conditions statisticians should propose the best method to analyze data sets. I have participated in the algorithm subgroup of Japan on the TRMM (Tropical Rainfall Measuring Mission) as a statistician. The following is my experience, which might be helpful for understanding cooperative studies between meteorologists and statisticians.

The TRMM satellite is planned to be launched by a Japanese H-II rocket in summer in 1997. One of the scientific objectives of TRMM is to estimate monthly mean rainfall over $5° \times 5°$ (latitude-longitude) boxes with a sampling error of less than 10%. Note that in this case the scientific goal is definite. The standard estimates will be provided by using the so-called deterministic method and the multiple-threshold method. The simple threshold method is considered as an effective method under study. Chiu (1988) empirically showed that there is a high correlation between the area-average rain rate and the fractional area that rain rate exceeds a fixed threshold value by using the GATE [GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment] data set. This fact provides the basis of the simple threshold method. Some meteorologists and statisticians published their cooperative studies on the threshold method. Among others are Kedem, Chiu and North (1990), Short, Shimizu and Kedem (1993) and Shimizu, Short and Kedem (1993). The mixed distribution as a statistical model was used to explain the reason why the threshold method works well. Under the model several methods for estimating higher moments and variance of rain rate have been proposed. At the beginning of the study there were some misunderstanding because of the significant difference between technical terms they use. From my experience the following two

- to know about the definite scientific goal of the study
- to know about properties of the data available

are extremely important factors for statisticians to suggest statistical methods and to build statistical models.

2. Statistical issues on spaceborne Lidar

2.1. Sparce sampling

The satellite altitude of the TRMM is low (350 km). Low-inclination (35°) TRMM orbit provides extensive sampling in the tropics. The satellite will revolve around the earth in about 90 minutes and make 16 revolutions a day and it will visit about twice daily at local times. Thus, the error due to sparce sampling is a critical issue. Oki and Sumi's (1994) paper studies sampling errors of monthly rainfall. Errors are estimated over 5° × 5° and 2.5° × 2.5° areas using a 43-month time series of radar-AMeDAS (Automatic Meteorological Data Aquisition System) data. Their study is unique because they used the orbital parameters for the real TRMM satellite and this procedure may be very informative even for the spaceborne Lidar experiment.

2.2. Missing values

There are many types of missing problems. It is known that the EM algorithm is an effective algorithm for getting numerical solutions (e.g., Little and Rubin, 1987). Theoretical solutions have been obtained only in limited cases.

2.2.1. Ozone measurements

Reinsel et al. (1994) performed a seasonal trend analysis of total ozone data. They showed some evidence of negative trend in the higher northern latitudes during the winter and spring seasons and also during all seasons.

Basu and Reinsel (1996) studied monthly averages of total stratospheric ozone measurements. Autoregressive moving average (ARMA) models may be used to model such kind of univariate time series if data are available at consecuitive and equally spaced time intervals. However, there are many cases where some of observations are missing. In fact, Basu and Reinsel used monthly averages of the ozone data at Huancayo, Peru (12°S, 75°W), which involve some entire months of missing observations. They used the restricted maximum likelihood (REML) estimation approach. A criticism of the usual maximum likelihood (ML) approach for the parameters of the error covariance structure in the presence of a regression component is that the ML approach does not take into account the loss in the degrees of freedom that results from estimating the regression parameters. The REML estimation approach overcomes this deficiency. A Kalman filter approach is another excellent estimation approarch. Data sets from the spaceborne Lidar may include some missing observations. It may be possible to use these estimation approaches.

2.2.2. Cloud data sets

In the book "Lognormal distributions" edited by Crow and Shimizu (1988), some description on cloud data sets can be seen in Chapter 13:

- The distribution of convective features accompanying a deep cumulonimbus cluster is as lognormal, "ranging from moderate cumulonimbus down to tiny, non-precipitating cumulus."
- Heights, horizontal sizes, and durations are lognormally distributed.

Shimizu (1993) proposed a bivariate mixed lognormal distribution to analize rainfall data measured at two stations. Zeroes, no rain, were treated as if the observations were missing. This model might be useful for analyzing cloud data sets. However, very recently Dr. Hayasaka of Tohoku University personally communicated a more serious problem to me. In my understanding there are high level, middle level, and low level clouds, and we can see the bottom from the ground (upper level clouds include missing data) and the top from the satellite (lower level clouds include missing data). An extension to the trivariate case of the bivariate mixed lognormal distribution seems useful as a statistical model for analyzing some cloud measurements, but estimation theory is open and it should be pursued in the future.

References

1. Basu, S. and G.C. Reinsel, 1996: Relationship between missing data likelihoods and complete data restricted likelihoods for regression time series models: an application to total ozone data. Appl. Statist., 45, 63-72.

2. Chiu, L.S., 1988: Rain estimation from satellites: Areal rainfall-rain area relations. 3rd Conference on Satellite Meteorology and Oceanography, American Meteorological Society, Jan. 31-Feb. 4, Anaheim.

3. Crow, E.L. and K. Shimizu, eds., 1988: Lognormal Distributions: Theory and Applications. Marcel Dekker, Inc.

4. Kedem, B., L.S. Chiu and G.R. North, 1990: Estimation of mean rain rate: Application to satellite observations. J. Geophys. Res., 95, 1965-1972.

5. Little, R.J.A. and D.B. Rubin, 1987: Statistical Analysis with Missing Data. John Wiley and Sons, Inc.

6. Oki, R. and A. Sumi, 1994: Sampling simulation of TRMM rainfall estimation using radar-AMeDAS composites. J. Appl. Meteor., 33, 1597-1608.

7. Reinsel, G.C.,G.C. Tiao, D.J. Wuebbles, J.B. Kerr, A.J. Miller, R.M. Nagatani, L. Bishop and L.H. Ying, 1994: Seasonal trend analysis of published ground-based and TOMS total ozone data through 1991. J. Geophys. Res., 99, 5449-5464.

8. Shimizu, K., 1993: A bivariate mixed lognormal distribution with an analysis of rainfall data. J. Appl. Meteor., 32, 161-171.

9. Shimizu, K., D.A. Short and B. Kedem, 1993: Single- and double-threshold methods for estimating the variance of area rain rate. J. Meteor. Soc. Japan, 71, 673-683.

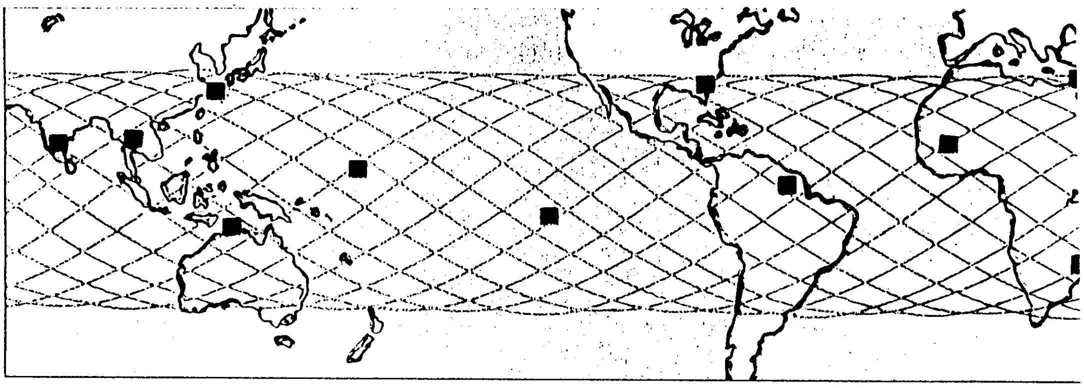
10. Short, D.A., K. Shimizu and B. Kedem, 1993: Optimal thresholds for the estimation of area rain-rate moments by the threshold method. J. Appl. Meteor., 32, 182-192.

Observation Strategy

Kunio Shimizu
Department of Applied Mathematics
Faculty of Science
Science University of Tokyo
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162

1. Statistician's point of view

2. Statistical issues on spaceborne Lidar
  2.1. Sparce sampling
  2.2. Missing values



Launch Vehicle : H-II Rocket
Altitude : 350 Km
Inclination : 35°

Example (TRMM)

- Chiu's (1988) empirical fact
- Estimation of area-average rain rate (basis of the threshold method)

$\Downarrow$

- Kedem, Chiu and North's (1990) explanation by using a statistical model

$\Downarrow$

- Short, Shimizu and Kedem's (1993) and Shimizu, Short and Kedem's (1993) estimation methods of higher moments

$\Downarrow$

- Discrimination of rain types

- Empirical study
- Looking at data

$\Downarrow$ This is a very important step for statisticians.

- Statistical model building

$\Downarrow$

- Development of statistical methods

$\Downarrow$

- Application
- Interpretation of the results

Two important factors for statisticians

- to know about the definite scientific goal of the study

- to know about properties of the data available

Some specific issues

1. Sparce sampling

2. Missing values
2.1. Ozone measurements

Reinsel et al. (1994) performed a seasonal trend analysis of total ozone data. They showed some evidence of negative trend in the higher northern latitudes during the winter and spring seasons and also during all seasons. **(Complete data)**

Basu and Reinsel (1996) studied monthly averages of total stratospheric ozone measurements. **(Incomplete data)**

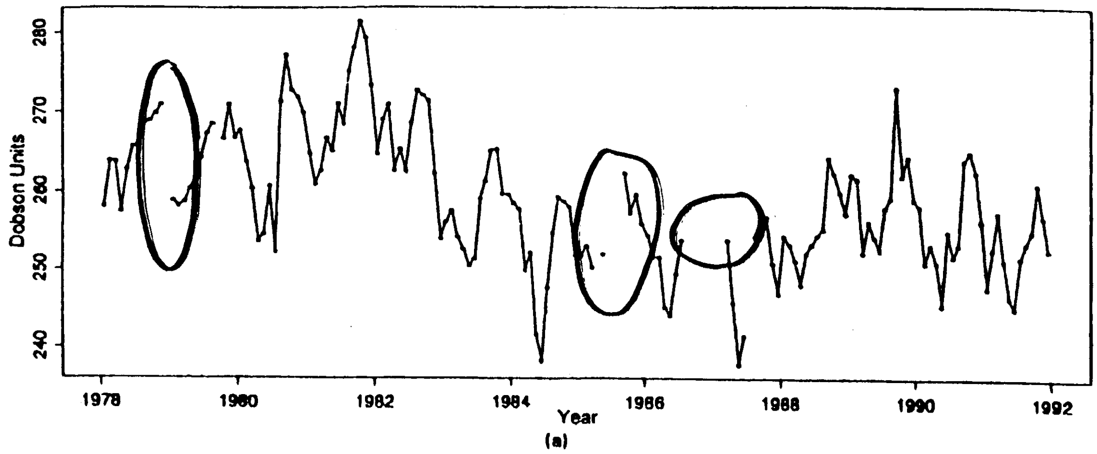


Fig. 1. (a) Monthly averages of total ozone from Huancayo, January 1978–December 1991

2.2. Cloud data sets

• The distribution of convective features accompanying a deep cumulonimbus cluster is as lognormal, "ranging from moderate cumulonimbus down to tiny, non-precipitating cumulus."

• Heights, horizontal sizes, and durations are lognormally distributed. (Crow and Shimizu eds., 1988, Chap. 13)

• Use of bivariate (Shimizu, 1993) and trivariate (apparently new!) mixed lognormal distributions