## 9. Statistical analysis of long series rainfall data: A regional study in southeast Asia.

Jean Pierre Bricquet[1], Arthorn Boonsaner[2], Ty Phommassack[3], Tran Duc Toan[4]

[1] IWMI SEA, Bangkok, Thailand
[2] Royal Forest Department, Bangkok, Thailand
[3] National Agronomy and Forestry Research Institute, Vientiane, Lao PDR
[4] National Institute for Soils and fertilizers, Hanoi, Vietnam
*Email address corresponding author :* j.bricquet@cgiar.org

**Abstract**
The occurrence of many extreme events in hydrology cannot be forecasted on the basis of deterministic information with the sufficient accuracy and lead time for those decisions which are sensitive to their occurrence. In such cases, a probabilistic approach is required in order to incorporate the effects of such phenomena into decision making.

A statistical study was conducted on long-term series data of daily rainfall from three weather stations in Vietnam (Hoa Binh), Laos (Luang Prabang) and Thailand (Phrae). These stations were all in relative close proximity to the MSEC (Management of Soil Erosion Consortium) study sites. Recorded periods ranged from 28 to 50 years. Analysis was undertaken on annual and maximum daily rainfall. Statistical adjustments derived values for decennial and centennial return period. Annual rainfall followed a normal distribution while maximum daily rainfall followed a Pearson 3 distribution.

All stations had reached annual values close to the millennium frequency (frequency 0.99) for the maximum values recorded and close to the centennial frequency (0.01) for the minimum values. All stations had reached the maximum daily values equal to or greater than the centennial value. The highest observed value for Hoa Binh station (416.4 mm) is probably a millennium occurrence. There was a small tendency for rainfall values to decrease even in 1997 which was a strong El Nino year in the region. From this study, we may adjust rainfall data inputs for the erosion model developed by MSEC and extend the study to a larger area.

**Introduction**
The occurrence of many extreme events in hydrology cannot be forecasted on the basis of deterministic information with sufficient accuracy and lead time for those decisions which are sensitive to their occurrence to be made. In such cases, a probabilistic approach is required to incorporate the effects of such phenomena into decisions. If the occurrence can be assumed to be independent of time, then frequency analysis can be used to describe the likelihood of any one or combination of events over the time horizon of a decision occurring (WMO, 1983).

Interpretation of precipitation has two major purposes. Firstly, to evaluate observations that sample a precipitation event or series of events. The evaluation of the observed sample includes consideration of extraneous influences, such as deficient or changing gauge exposure, and interpretation of the effects of physical environment, such as physiography. Secondly, to describe the event in a form appropriate for display, subsequent analysis, or other applications.

There is a paucity of information related to trends in rainfall in the Southeast Asia region. A study conducted by Manton *et al.* (2001) showed a gap in the Indochinese peninsula. Through the Management of Soil Erosion Consortium (MSEC) led by the International Water Management Institute (IWMI), it was possible to gain access to three recorded series in this region. This paper provides a statistical analysis of long term rainfall data in the region.

**Location of the study**

The analysis was done using the data from three weather stations relatively close to the MSEC project sites. These were Hoa Binh in Vietnam (20° 49' N, 105° 20' E), Luang Prabang in Lao PDR (19° 53' N, 102° 08' E), and Phrae in Northern Thailand (18° 08'N, 100° 10'E) (Figure 1). The period of record is 41 years for Hoa Binh, 51 for Luang Prabang and 28 for Phrae. Daily rainfall data are available at all stations.


**Figure 1.** Location of the three rainfall recording stations (★) in Thailand, Laos and Vietnam

**Results**

*Homogeneity of the Data*

The homogeneity of hydrological or meteorological data is a primary requirement for valid statistical applications. The most commonly used information about non-climatic influences comes from records of station movement, changes in instrumentation, problems with instrumentation, sensor calibration, changes in surrounding environmental characteristics, observation practices, and other similar features (Guttman, 1998). The double mass curve analysis introduced by Kohler (1949) is a graphical method of identifying and adjusting inconsistencies in a station record by comparing its time trend with those of other stations. Changes in slope of double-mass curves may be caused by changes in exposure or location of gauge, changes in procedure in collecting and processing data, etc (WMO, 1994). The data collected from all the sites within the region should be highly correlated, have similar variability, and differ only by scaling factors and random sampling variability.

As shown in Figure 2, there is no significant change in the slope of the individual curves for both Hoa Binh and Phrae stations when plotted against the accumulative rainfall of Luang Prabang. So, we can consider all series as homogenous. With the highly significant correlation coefficients, it is also possible to extend the series of Hoa Binh and Phrae but only for annual or monthly values.

*Annual Rainfall Analysis*

Normal Climate is defined by the World Meteorological Organization (1983) as "period averages of a climatic element such as temperature or precipitation computed for a uniform and relatively long period comprising at least three consecutive ten-year periods". Manton *et al*, (2001) assumes that the annual total rainfall has generally decreased between 1961 and 1998 and the number of rainy days (with at least 2 mm of rain) has decreased significantly in Southeast Asia. Reviewing the variation in annual total rainfall (Figure 3) of the three stations, it is difficult to detect a trend of either an increase or decrease annual rain.

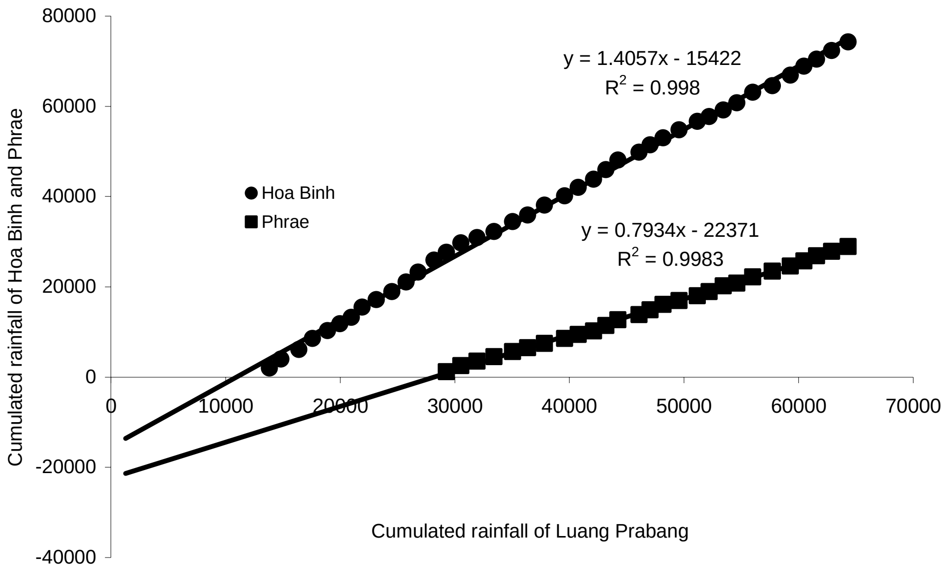


**Figure 2.** Double mass curves for Hoa Binh and Phrae stations with respect to the cumulative rainfall of Luang Prabang.

.

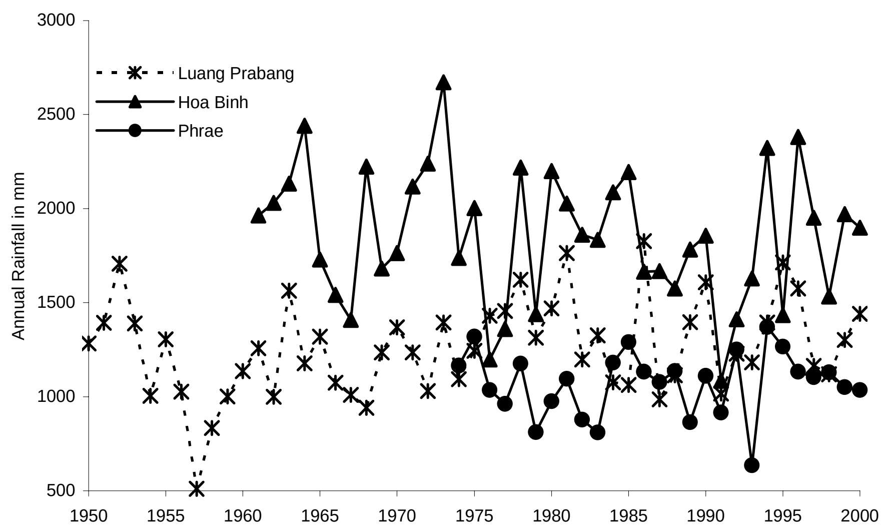


**Figure 3.** Variation in the annual rainfall at the three recording stations within the study.

The 1957 value of Luang Prabang appear to be an aberration. However, checking the records, we can not reject this very small amount of rain (511.1 mm). Similarly the same would apply for the 1993 value for Phrae (635.9 mm). Descriptive statistics for each station are presented in Table 1.

**Table 1**. Summary statistics of the annual rainfall of the three stations.

| | **Average** | **Variance** | **Std. deviation** | **Minimum** | **maximum** | **Std. kurtosis** |
|---|---|---|---|---|---|---|
| Hoa Binh | 1856.7 | 127 290 | 356.8 | 1085.2 | 2671.7 | -0.473 |
| Luang Pranbang | 1261.2 | 65 058 | 255.1 | 511.1 | 1827.5 | 0.817 |
| Phrae | 1084.6 | 34 107 | 184.7 | 635.9 | 1461.3 | 0.273 |

Annual rainfall usually follows a Gaussian distribution (Figures 4). Accordingly, the different return period values with the Normal law were calculated and are presented in Table 2.

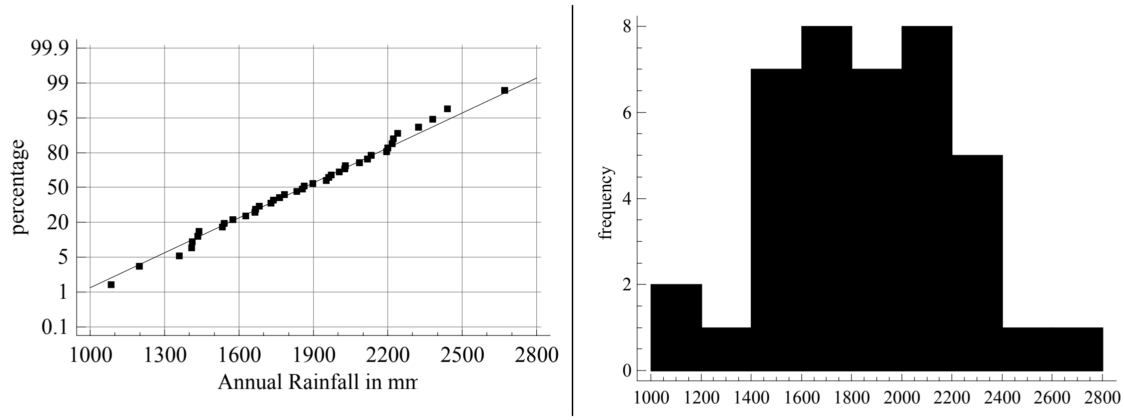


**Figure 4.** Annual rainfall at Hoa Binh: distribution plotting

**Table 2.** Frequency analysis of the annual rainfall of the three stations.

| | 0.01 | 0.1 | 0.5 | 0.9 | 0.99 |
|---|---|---|---|---|---|
| Hoa Binh | 1064.2 | 1420.8 | 1856.7 | 2292.7 | 2649.3 |
| Luang Pranbang | 710.7 | 958.4 | 1261.2 | 1563.9 | 1811.6 |
| Phrae | 654.9 | 848.3 | 1084.6 | 1321.0 | 1514.3 |

According to the analysis of these data sets it is clearly evident that all stations had already reached annual values close to the millennium frequency (frequency 0.99) for the maximum values recorded. Similarly, for minimum values, all stations have already reached centennial frequency (0.01).

*Maximum Daily Rainfall*

Figure 5 presents the frequency distribution of maximum daily rainfall for the three stations. For this type of data, we don't have a Gauss distribution and the data can be adjusted with a Log Normal distribution (Figure 6). Descriptive statistics for maximum daily rainfall are presented in Table 3.

**Table 3.** Summary statistics of the maximum daily rainfall of the three stations.

| | Average | Variance | Std. Deviation | minimum | maximum | Std. kurtosis |
|---|---|---|---|---|---|---|
| Hoa Binh | 156.1 | 4 293.2 | 65.5 | 58.3 | 416.4 | 6.620 |
| Luang Pranbang | 83.2 | 852.8 | 29.2 | 27.7 | 180.7 | 3.615 |
| Phrae | 90.2 | 1 278.8 | 35.7 | 43.3 | 218.2 | 7.062 |
| | | | | | | |

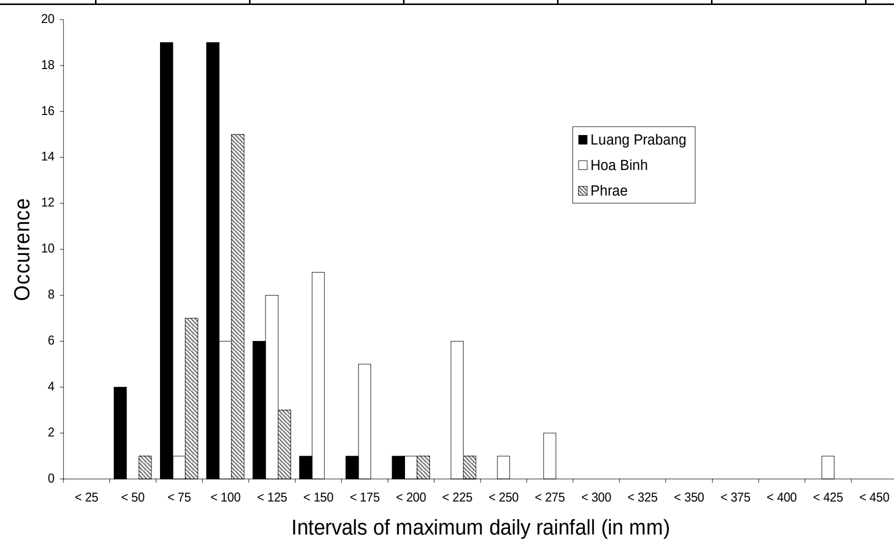


**Figure 5.** Frequency histogram of maximum daily rainfall at the three stations.

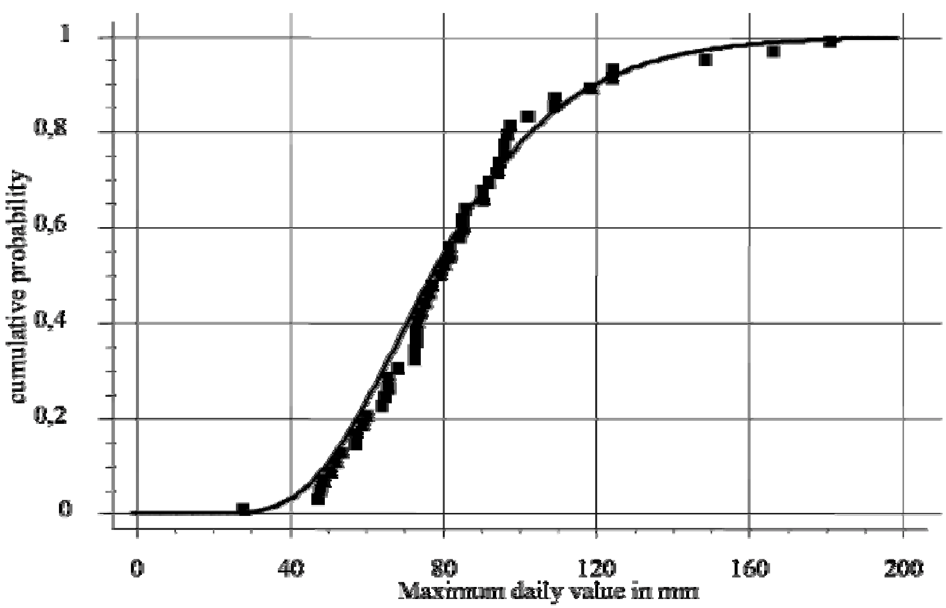


**Figure 6.** Adjustment of maximum daily rainfall at Luang Prabang.

The calculation of return period values were undertaken following a Pearson 3 law and the results are presented in Table 4.

**Table 4.** Frequency analysis of the maximum daily rainfall of the three stations.

| | 0.01 | 0.1 | 0.5 | 0.9 | 0.99 |
|---|---|---|---|---|---|
| Hoa Binh | 64.4 | 90.2 | 142.6 | 238.6 | 378.9 |
| Luang Pranbang | 34.7 | 50.7 | 79.2 | 121.0 | 168.1 |
| Phrae | 48.4 | 58.0 | 81.3 | 132.4 | 222.7 |

The highest observed value for Hoa Binh station (416.4 mm) is probably a millennium occurrence. It is also clearly evident that all stations have reached maximum daily values equal to or higher than the centennial calculation. The same comment can be ascribed to the lower values.

## Conclusion

This study is the prerequisite for a more in depth investigation into the climate of the region. As we can detect a decrease of the annual total rainfall for the three stations, we should continue this study with more stations from the region including South China. A limitation of such studies is the low spatial density of stations with homogenous data.

## Acknowledgment

This study is a contribution to the Management of Soil Erosion Consortium led by the International Water Management Institute and funded by the Asian Development Bank (RETA 5803). Special thanks to the National Departments of Meteorology of the different countries which provided the data.

## References

1. Guttman N.B., 1998: Homogeneity, data adjustments and climatic normals, 7th International Meeting on Statistical Climatology, Whistler, Canada, May 25-29, 1998. In Abstract Booklet.
2. Karl T.R. and Knight R.W., 1998: Secular trends of precipitation amount, frequency, and intensity in the United States, *Bull. Amer. Meteor. Soc.*, **79**, 1107-1119.
3. Kohler M. A., 1949: Double-mass analysis for testing the consistency of records for making adjustments. *Bull. Amer. Meteor. Soc.*, **30**, 188-189
4. Manton M.J., Della-Marta P.M., Haylock M.R., Hennessy K.J., Nicholls N., Chambers L.E., Collins D.A., Daw G., Finet A., Gunawan D., Inape K., Isobe H., Kestin T.S., Lefale P., Leyu C.H., Lin T., Maitrepierre L., Outprasitwong N., Page C.M., Pahalad J., Plummer N., Salinger M.J., Suppiah R., Tran V.L., Trewin B., Tibig I., and Yee D., 2001: Trends in extreme daily rainfall and temperature in southeast Asia and the South Pacific, *Int. J. Climatol.* **21**, 269-284.
5. World Meteorological Organization, 1994: Guide to hydrological practices, WMO-No 168.
6. World Meteorological Organization, 1989: Calculation of monthly and annual 30-year standard normal. WMO-TD/No. 341. Geneva. 11 p.
7. World Meteorological Organization, 1983: Guide to climatological practices, WMO-No 100.