Nordic Hydrology, 23, 1992, 183-192

Linear or Nonlinear Covariance of Seasonal Snowmelt and Snow Cover in Western Himalayas

B. Dey and V. K. Sharma
Howard University, Washington, DC 20059

A. Rango
USDA Hydrology Lab., Beltsville, MD 20705

Log-linear, exponential and fractional relations for estimating seasonal snowmelt from early-spring snow accumulation in the Indus and Kabul river basins in the western Himalayas are developed with a view to improve the prediction given by bivariate linear regression models earlier developed by the senior author in collaboration with others. This study shows that although the transformed data may improve the above prediction, they fail to satisfy the condition of nonlinearity; a property that must be borne in mind before recommending any nonlinear regression model. Any further improvement in the prediction of seasonal flow volume from basin snow cover area, therefore, has to come from within the domain of linear regression models only or from improvements in the original input data.

Introduction

Previous investigations have suggested a linear correlation between the summer seasonal snowmelt discharge and early-April snow cover in some of the Himalayan watersheds (Rango *et al.* 1977; Tarar 1982; Dey *et al.* 1983). Linear regression models assume that basin discharge, the dependent variable, increases by a constant amount with a unit increase in the value of basin snow cover area, the independent variable. A relationship of this type can be completely specified by mean values, variance and covariance of the two variables only (Rao 1965). In natural phenomena, however, this assumption of linearity may not always be satisfied either due to the effect of saturation or the inherent nature of the independent

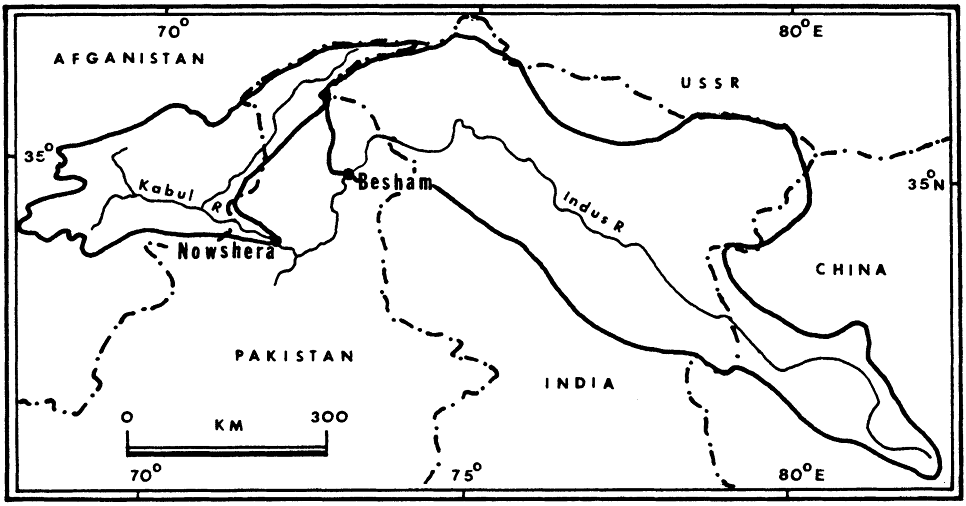

Fig. 1. Indus and Kabul basins.

variable which eventually imposes limits on the dependent variable. Gupta *et al.* (1982), for example, have shown that the relationship between the meltwater yield and snow cover area in the Beas watershed and its subbasins in the western Himalayas is exponential. Suspecting that similar relations may be obtained also for the Indus and Kabul basins (Fig. 1), three transformed linear regressions based on nonlinear assumptions were developed to describe the covariance from 1969-79 of April through July flow volume and April 1-20 per cent snow cover area for the same period. It was further held that one of these models will improve the prediction of seasonal flow volume when compared to the bivariate regression flow models for the two rivers developed earlier by Dey *et al.* (1983).

The Indus River originates at an elevation of 5,182 m on the Tibetan plateau near Mansrover Lake in China. From its source to Besham, the gauging station at 1,200 m in northern Pakistan, the river drains 162,000 km^2 area, most of which is situated in India. The major tributaries of the Indus have their catchments in the lofty northwest-southeast trending Karakoram, Ladhak, Zaskar, and Great Himalayan ranges where they are fed by meltwater from snow fields and glaciers of large dimensions. In this region of highly deformed rocks, Mt. Godwin Austin in the Karakoram range forms the highest peak (8,611m), and it is followed in elevation by the Naga Parbat (8,126m) in the Great Himalayan range. The climate of the catchment in the Tibetan plateau is cold-dry, and elsewhere in the perpetually snow covered peaks and slopes it is alpine in nature. The Indus catchment in India receives an estimated 56 cm of mean annual precipitation. The meltwater yield of the catchment for April through July months of 1969 to 1979 has varied between 20 and 35 cm depth of runoff.

The Kabul River drains 88,600 km^2 in Afganiștan and Pakistan. The elevation in the basin varies from 7,690 m in the valley of Konar, the principal tributary of Kabul, to 305 m at Nowshera, the gauging station in Pakistan. The summer season flow in the Kabul is determined primarily by snowmelt in a period of high thermal efficiency. The April through July runoff depth varies from 12 to 25 cm in the catchment for the eleven years of streamflow record analyzed in this research.

Methodology

The 1969-79 April 1-20 snow cover area as estimated from NOAA/TIROS images, and the April through July observed streamflows for the same years in the Indus above Besham and Kabul near Nowshera were tabulated by Dey *et al.* (1983). For these data, three regression models assuming a) log-linear covariance of discharge and snow cover area, b) exponential variation of discharge with snow cover area, and c) bivariate relation of the square-root of discharge and basin snow cover area are developed (Fig. 2) to provide a rational functional explanation of seasonal snowmelt phenomenon from the basin snow cover area. In the log-linear or power function of the type $Y = ax$, the covariance of both snow cover area and seasonal snowmelt is in geometric progression and governed by a property known as constant elasticity. The plot of these points on double logarithmic paper tends to be in a linear pattern. The exponential form $y = ab^x$ represents the logarithm of seasonal runoff as a linear function of snow cover area, the plots of which on semi-logarithmic scale show a constant ratio or proportion of change in amount of change as a straight line. The functional relationship of the type such as, $\sqrt{y} = a+bx$, is based on the assumption that the plot of square-root of meltwater yield as predictor variable and the explanatory snow cover parameter define a linear pattern of distribution. These equations are analyzed for the diagnostic statistics and the tests of significance bearing on the bivariate relationships.

Results and Discussion

The three transformed linear regressions for the Indus and Kabul basins are shown in Fig. 2. On the bases of coefficient of determination and standard error of estimate computed in Table 1, an exponential relation for the Indus and log-linear function for the Kabul watershed appear to best describe the covariance of seasonal snowmelt yield with the respective basin's snow cover extent. These relationships show slight improvement in the prediction of seasonal runoff volume from snow cover area given by the linear regression models developed by *Dey et al.* (1983) for the two watersheds. Acceptance of these two transformed linear regressions would, however, require explanation of the functional relationship between

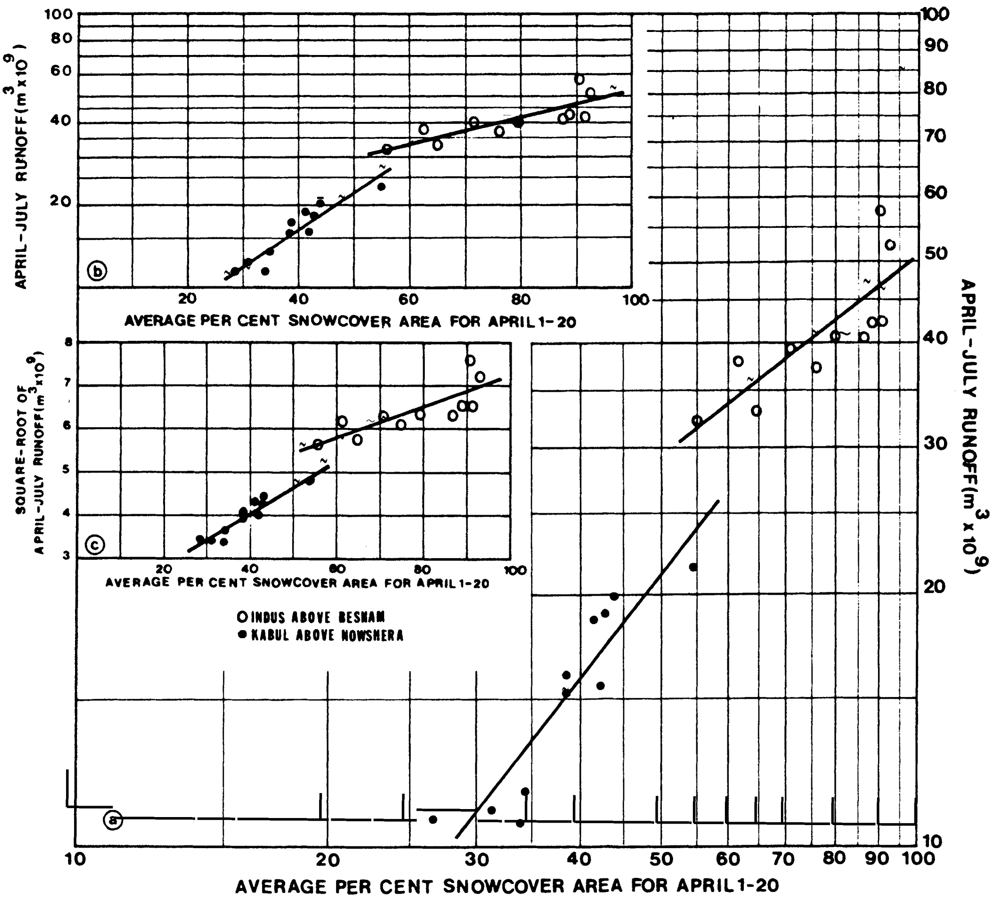

Fig. 2. Transformed linear regression on nonlinear assumptions.

snowmelt and snow cover area by two hypotheses. Such a proposition does not appear to be consistent with reality since the Himalayan watersheds reportedly are characteristic of innate concurrent flow conditions during the melt season (Dey and Goswami 1984).

It was, therefore, decided to examine the significance of transformed sample data to the overall regression. The standard error of regression slope (S_b) is a diagnostic statistical parameter that describes the significance of independent variable to the explanation of dependent variable in a regression model. It may be expressed as

$$S_b = \frac{S_y}{\sqrt{\Sigma (x_i - \bar{x})^2}} \qquad (1)$$

where, S_y is the standard error of estimate provided by the given regression model

Table 1 – Comparison of transformed linear regressions on nonlinear relations, and linear regression models of seasonal snowmelt y and percent snow cover area x in the Indus and Kabul basins, 1969-79

Relationship	Indus above Besham	Kabul above Nowshera
a. Log-linear: log y, log x		
Estimating equation:	$y=1.3224\ x^{0.7903}$	$y=0.1584\ x^{1.2616}$
Coefficient of determination:	0.6356	0.8832
Standard error of estimate:	$0.0455\times10^9 m^3$	$0.0357\times10^9 m^3$
b. Exponential: log y, x		
Estimating equation:	$\log y=0.0047x+1.2446$	$\log y=0.0137x+0.6652$
Coefficient of determination:	0.6540	0.8705
Standard error of estimate:	$0.0443\times10^9 m^3$	$0.0375\times10^9 m^3$
c. Fractional: $\sqrt{y}$, x		
Estimating equation:	$\sqrt{y}=0.0350x+3.6903$	$\sqrt{y}=0.0636x+1.5223$
Coefficient of determination:	0.6245	0.8888
Standard error of estimate:	$0.3512\times10^9 m^3$	$0.1601\times10^9 m^3$
d. Linear*: y, x		
Estimating equation:	$y=0.4536x+6.1453$	$y=0.5205x-4.07$
Coefficient of determination:	0.60	0.90
Standard error of estimate:	$4.63\times10^9 m^3$	$1.17\times10^9 m^3$

* Source: Dey et al. (1983)

and the denominator the summation of squared differences in the snow cover area from its mean of eleven observations. The standard error of regression slope may be interpreted in a manner similar to the standard deviation.

The significance of the relationship between the independent and dependent variable may be tested by setting-up a null hypothesis that the true slope $\equiv 0$ and an alternate hypothesis that the slope $\neq 0$. The estimated value of regression slope, b, may be converted to the t-scale by

$$t \equiv \frac{b}{S_b} \tag{2}$$

The null hypothesis, if accepted at n-2 d.f. at 95 per cent level, leads to the conclusion that the independent variable is not useful in explaining the dependent variable. The t-test for the significance of regression slope given in Table 2 reveals that the data transformed for analysis by linear regressions assume such characteristics that, despite giving a higher coefficient of determination in some models, the extent of basin snow cover area does not explain the flow volume during the snowmelt season. This is the case for the log-linear covariance in both the Indus and Kabul basins.

Table 2 – The t-test for the significance of regression slope

Relation	S_b	t	Tabulated t at 9 d.f. and 95 percent level	Remarks about null hypothesis
		A. Indus above Besham		
Log-linear	0.0291	1.156	2.262	Accept
Exponential	0.0018	2.557	2.262	Reject
Fractional	0.0140	2.498	2.262	Reject
		B. Kabul above Nowshera		
Log-linear	0.0565	1.745	2.262	Accept
Exponential	0.0046	2.950	2.262	Reject
Fractional	0.0213	2.981	2.262	Reject

Having concluded that the independent variate in transformed sample data does not significantly explain the dependent variate for the log-linear transformation, it was, for confirmation, decided to analyze whether a straight line is to be preferred over a given nonlinear relation. This method is based on the analysis of variance in which the extent to which a given nonlinear relation reduces the residuals about the straight line is compared using the F ratio (see Rosander 1957, Ch. 32). The analysis for the Indus and Kabul basins summarized in Table 3 suggests that for all the three nonlinear relations, the improvement over linear regression is not significant. Hence, transformed linear regressions on nonlinear relations are ruled out for the two watersheds.

Having established that the three nonlinear regressions developed in this research are not significant improvements over the linear regression for the prediction of seasonal snowmelt yield, it was decided to test whether any other form of curvilinear regression could be suitably considered for application to the data. For this purpose, correlation ratio, η^2, is a suitable measure. If η^2 is significantly larger than the coefficient of determination, r^2, the regression is given to be curvilinear (see Rosander 1957, Chs. 22 and 32). The significance of the difference between η^2 and r^2 is tested by the following expression for the F test

$$F \equiv \frac{\eta^2 - r^2}{k-2} \; \frac{\eta - k}{1-\eta^2} \qquad (3)$$

in which n is the size of sample and k the number of columns.

* F test is not applicable as the log-linear relation increases the residuals about the linear regression.

\+ Tabulated $F_{1,8}$ at 5% $\equiv$ 5.32.

Table 3

Table 3 – Analysis of variance for the residuals

A. Indus above Besham

Component	Degrees of Freedom	Sum of Squares	Mean Square	*F*
Linear function	1	325.27	325.27	–
Log-linear function about linear relation	1	– 4.33*	–	–
Residual variation	8	240.34	–	–
Total	10	561.28		
Linear function	1	325.27	325.27	11.35
Exponential function about linear relation	1	6.77	6.77	0.24+
Residual variation	8	229.24	28.65	
Total	10	561.28		
Linear function	1	325.27	325.27	11.23
Fractional function about linear relation	1	4.27	4.27	0.15+
Residual variation	8	231.74	28.96	
Total	10	561.28		

B. Kabul above Nowshera

Component	Degrees of Freedom	Sum of Squares	Mean Square	*F*
Linear function	1	137.10	137.10	–
Log-linear function about linear relation	1	– 0.48*	–	–
Residual variation	8	15.53	–	–
Total	10	152.15		
Linear function	1	137.10	137.10	52.73
Exponential function about linear relation	1	5.10	5.10	1.96+
Residual variation	8	20.57	2.6	
Total	10	152.15		
Linear function	1	137.10	137.10	66.55
Fractional function about linear relation	1	1.45	1.45	0.70+
Residual variation	8	16.50	2.06	
Total	10	152.15		

Table 4 – Summary statement of the significance of correlation ratio

	Parameter			
Basin	r^2	η^2	Computed F	Tabulated F for the given degrees of freedom at 5%
Indus	0.5418	0.5981	0.4902	$F_{2,7}$ = 4.74
Kabul	0.8704	0.8854	0.1631	$F_{4,5}$ = 5.19

The summary statistics of correlation ratio is given in Table 4. As the computed F is lower than the tabulated F for the given degrees of freedom at 5 per cent, it may be inferred that curvilinear regressions do not add significantly to the linear regression, and that only linear regressions are recommended for both the Indus and Kabul basins.

Despite improving the standard error of estimate and coefficient of determination, the covariance between snowmelt and snow cover area is not explained by any one single nonlinear regression model. Furthermore, the sample data also do not satisfy the statistical conditions of nonlinearity and curvilinear distributions in the Indus and Kabul basins. Therefore, the functional relationship between snowmelt and snow cover area is explained only by the linear rather than nonlinear regression models.

Prediction of seasonal snowmelt by linear regression models may be improved by extending the period of satellite-observed snow cover area and the corresponding streamflow record. Whereas it is further possible to improve the manually-processed satellite snow cover area by using computer-associated digital processing of snow cover area (*e.g.* Carroll 1990), the quality of available streamflow data however will remain suspect.

As more detailed observations of temperature and precipitation become available in these remote regions, the prediction of snowmelt runoff could be improved by using a distributed hydrological model similar to the one developed by Rango and Martinec (1979). Forecasts using a snowmelt-runoff model require input of estimates of future snow covered area. Such estimate are available by using the so-called modified depletion curves of snow coverage (Martinec and Rango 1987). Such a deterministic approach would be preferable to the empirical approaches used here.

A major limiting factor in this study and other studies in developing countries is a lack of an adequate data base. Improved and longer periods of hydrological and meteorological data would allow a better understanding of point and areal meteorological conditions during snow accumulation and melt periods, the use of deterministic instead of empirical methods to estimate water yield, improved statistical relationships with confidence in conclusions derived from the statistical analysis, and the possibility of evaluating the accuracy of data used in analysis such as snow cover and runoff. Unfortunately, extension of the data base in terms of spatial and temporal scales is a monumental task. Remote sensing data are improving the situation. However, one will have to wait for some more time before remote sensing information will be available over long time periods. Until such time, conclusions from studies in these important areas of the world will have to be made as best possible using the data currently available. The results from this study were made under the limitations outlined above.

References

Carroll, T. R. (1990) Airborne and satellite data used to map snow cover operationally in the U. S. and Canada, Proceedings of the International Symposium on Remote Sensing and Water Resources, Enschede, The Netherlands, pp. 147-155.

Dey, B., and Goswami, D. C. (1984) Evaluating a model of snow cover area versus runoff against a concurrent flow correlation model in the western Himalayas, *Nordic Hydrology, Vol. 15*, pp. 103-110.

Dey, B., Goswami, D. C. and Rango, A. (1983) Utilization of satellite snow cover observations for seasonal streamflow estimate in the western Himalayas, *Nordic Hydrology, Vol. 14,* pp. 257-266.

Gupta, R. P., Duggal, A. J., Rao, S. N., Sankar, G., and Singhal, B. B. S. (1982) Snow-cover area *vs*. snowmelt runoff relation and its dependence on geomorphology – a study from the Beas catchment (Himalayas, India), *Jn. of Hydrology, Vol. 58,* pp. 325-339.

Martinec, J., and Rango, A. (1987) Interpretation and utilization of areal snow cover data from satellites, *Annals of Glaciology, Vol. 9,* pp 166-169.

Rango, A., and Martinec, J. (1979) Application of a snowmelt-runoff model using Landsat data, *Nordic Hydrology, Vol. 10,* pp. 225-238.

Rango, A., Solomonson, V. V., and Foster, J. L. (1977) Seasonal stream flow estimation in the Himalayan region employing meteorological satellite snow cover observations, *Water Resources Research, Vol. 13,* pp. 109-112.

Rao, C. R. (1965) *Linear Statistical Inference and its Applications*, New York, John Wiley & Sons, Inc.

Rosander, A. C. (1957) *Elementary Principles of Statistics,* Toronto, D. Van Nostrand Co. Inc.

Tarar, R. N. (1982) Water resources investigations in Pakistan with the help of Landsat imagery – snow survey, 1975-78, Hydrological Aspects of Alpine and High Mountain Areas (Proceedings of the Exeter Symposium), IAHS Publication No. 138, pp. 177-190.

Received: 13 June, 1991
Revised version received: 10 February, 1992
Accepted: 14 April, 1992

Address:
B. Dey,
Department of Physics and Astronomy,
Howard University,
Washington, DC-20059,
U.S.A.

V. K. Sharma,
Flats D-6,
University Campus,
Kurukshetra,
Haryana-132119,
India.

A. Rango,
Hydrology Laboratory,
USDA,
Beltsville, MD-20705,
U.S.A.